

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 11:18:53 ; Search time 32.5 Seconds
(without alignments)
1233.774 Million cell updates/sec

Title: US-09-839-894-10

Perfect score: 1886
Sequence: 1 MKKILFIFTLFFSSVLFTFA.....EAGQYMGNIWTFPPSSQTL 361

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : A.GeneSeq.032802.*
1: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneSeq/geneSeq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	927.5	49.2	364	AAV22326	Pilin protein CotD
2	275.5	14.6	359	AA45919	S. enterica serova
3	119.5	6.3	770	AAV50814	D. melanogaster ac
4	114	6.0	15	AA47575	Alpha-2-glycyltin o
5	111.5	5.9	1005	AA801833	Haemophilus influe
6	111.5	5.9	1011	AA801832	Haemophilus influe
7	110.5	5.9	1529	AA41732	High molecular wei
8	110.5	5.9	1601	AAV30292	Non-typeable Haemo
9	109	5.8	2383	AA815945	E. coli proliferat
10	107	5.7	1095	AA801835	Haemophilus influe
11	107	5.7	1101	AA801834	Haemophilus influe

12	106	5.6	671	AA85290	Streptococcus faec
13	106	5.6	1222	AA801830	H. influenzae stra
14	106	5.6	1228	AA801828	Haemophilus influe
15	101	5.4	1338	AA41731	High molecular wei
16	101	5.4	1598	AAV30291	Non-typeable Haemo
17	100.5	5.3	1004	AA801841	Haemophilus influe
18	100.5	5.3	1010	AA801840	Haemophilus influe
19	99.5	5.3	969	AA801827	Haemophilus influe
20	99.5	5.3	975	AA801826	Haemophilus influe
21	99	5.2	992	AA801843	Haemophilus influe
22	99	5.2	998	AA801842	Haemophilus influe
23	97	5.1	418	AA830907	Arabidopsis thalia
24	97	5.1	454	AA830906	Arabidopsis thalia
25	97	5.1	957	AA801839	Haemophilus influe
26	97	5.1	963	AA801838	Haemophilus influe
27	96	5.1	450	AA837170	Arabidopsis thalia
28	96	5.1	513	AA839277	Arabidopsis thalia
29	96	5.1	631	AA839276	Arabidopsis thalia
30	96	5.1	683	AA839275	Arabidopsis thalia
31	95.5	5.1	825	AA825547	Eucalyptus grandis
32	94.5	5.0	567	AA837170	Amino acid sequenc
33	94.5	5.0	665	AA868244	Drosophila melanog
34	94.5	5.0	2353	AA893933	Haemophilus adhesi
35	94.5	5.0	2411	AA823860	Haemophilus influe
36	94	5.0	394	AA837668	Arabidopsis thalia
37	93.5	5.0	715	AAV05625	HIV-1 group O isol
38	93.5	5.0	1464	AA871111	Drosophila melanog
39	93	4.9	354	AAV55510	H. pylori ORF 04ge
40	93	4.9	444	AAV55696	H. pylori ORF 13ae
41	93	4.9	444	AAV17193	H. pylori outer me
42	93	4.9	808	AA858477	Drosophila melanog
43	93	4.9	1167	AAV31504	Nematox toxin 167
44	93	4.9	1167	AAV10653	Bacillus thuringie
45	92.5	4.9	1073	AA801837	Haemophilus influe

ALIGNMENTS

RESULT 1	AAV22326	standard; Protein: 364 AA.
ID	AAV22326	
XX	AAV22326;	
AC	22-SEP-1999	(first entry)
XX		
DE	Pilin protein CotD.	
XX		
KW	CS2 gene cluster; CotA; CotB; CotC; CotD; pilin protein; immunogen;	
KW	enterotoxigenic E. coli; human upper intestine; diarrhoeal disease;	
KW	enteric infection; therapy.	
XX		
OS	Escherichia coli.	
XX		
PN	US5932715-A.	
XX		
PD	03-AUG-1999.	
XX		
PF	07-JUN-1995;	95US-0483101.
XX		
PR	07-JUN-1995;	95US-0483101.
XX		
PA	(UYEM-) UNIV EMORY.	
XX		
PI	Caron J, Froehlich B, Scott JR;	
XX		
DR	WPI; 1999-443623/37.	
XX		
PT	N-PSDB; AA84848.	
PT	Isolated nucleic acids encoding Escherichia coli CS2 pilin proteins	
PT	useful for vaccinating against diarrhoeal diseases caused by	
PT	Escherichia coli	

KW neurotransmission; plant protection agent; conductance; AChR.
 XX Drosophila melanogaster.
 XX DE19819829-A1.
 XX 11-NOV-1999.
 XX 04-MAY-1998; 98DE-1019829.
 XX 04-MAY-1998; 98DE-1019829.
 XX (FARB) BAYER AG.
 XX Adamczewski M, Oellers N, Schulte T;
 XX WPI: 2000-014207/02.
 DR N-PSDB; AA24475.
 XX
 PT New nucleic acid encoding a nicotinic acetylcholine receptor from
 PT insects, used to identify potential insecticides -
 XX
 PS Example 1a; Page 12-14; 26pp; German.
 XX
 CC This invention describes a novel nucleic acid (NA) encoding a nicotinic
 CC acetylcholine receptor (R) from insects which can be used as an
 CC insecticide. Inhibitors of (R) interfere with neurotransmission. (R)
 CC (also vectors containing it, its regulatory regions, and antibodies
 CC directed against (R)-encoded proteins) are used to screen for: (a) plant
 CC protection agents that alter conductance of AChR, potentially useful as
 CC insecticides, or (b) genes which encode polypeptides that are involved in
 CC formation of functionally related AChR in insects. (R) are also used to
 CC isolate and characterize the specified regulatory regions and for
 CC recombinant production of (R). This sequence represents an
 CC acetylcholine receptor isolated from *Drosophila melanogaster*.
 CC
 XX Sequence 770 AA;
 XX

Query Match 6.3%; Score 119.5; DB 21; Length 770;
 Best Local Similarity 20.4%; Pred. No. 0.034;
 Matches 83; Conservative 48; Mismatches 140; Indels 135; Gaps 20;

QY TLFFSVLEFTEFAVSADKIPGDESTINIFGPRDRNESPKNILNNH----- 54
 DB 249 tlatfslvgstaaglkmsssssssn-----ssmstqlnglnkhswlfillynlsak 304
 QY 55 --ITAYSESHLYDRMFLCLSSHNTLNGACPTSENPSSSVSGETNTTLQFTEKRLIK 112
 DB 305 vclaglyhekrllhd-----lldpntlerpylnesdplqsfyltltmglidvdeknqllv 359
 QY 113 RELQIKGYKQLLFKSVNCPGLTNSAHFNCKNAASGLSYLYTPAGELNLPFGIWD 172
 DB 360 tnvajk-----lewdmnlrwtadsgyvk-----dlr-iphrtwk 395
 QY 173 ATLRLRVKRRYSER---YGTITINITIKLTDKGNIOQLWLO--FKSNARDNLNRPFGG 226
 DB 396 pdvilm-----ynsdegfdgtyqtnvvrr--nngsc-lyvppgalfkscckidltwfp--- 444
 QY 227 GTYIGRNSVDKCF---YDGSTNSSLEIRFODNNPKS-----DGKFLRKINDDKRE 276
 DB 445 ---fdqgcemkfgswcydgf-----qldlqldetgdisylngewellgyppgkxne 496
 QY 277 IAYT-----LSLLLAGKSLTPNTGSLNLNADA 303
 DB 497 lyncncepepyiditfaiirrtlyffnllipcvllasmallgftlpdpdsgeklsgvt 556
 QY 304 ASLE-TMNNRTAYTMPKIS--VPV-----LCW-----PGR 331
 DB 557 lllstlvllmvaetmpatsdaplwlrlvflcwlplmlmsrpr 602

RESULT 4

AA047575
 ID AA047575 standard; Protein; 650 AA.
 XX
 AC AA047575;
 XX
 DT 19-JUL-1994 (first entry)
 XX
 DE Alpha-agglutinin of *Saccharomyces cerevisiae*.
 XX
 KW Immobilisation; enzyme; cell wall; alpha agglutinin; AGA 1; FLO 1;
 KW Major cell wall protein; glycosyl-phosphatidylinositol;
 KW anchoring protein; alpha factor; alpha-agglutinin; invertase;
 KW inulinase; alpha-amylase; *Saccharomyces cerevisiae*;
 KW enzymatic process; fermentation; biodegradation; catalysis.
 XX
 OS *Saccharomyces cerevisiae*.
 XX
 PN W09401567-A.
 XX
 PD 20-JAN-1994.
 XX
 PF 07-JUL-1993; 93WO-EP01763.
 XX
 PR 08-JUL-1992; 92EP-0202080.
 PR 14-DEC-1992; 92EP-0203899.
 XX
 PA (UNIL) UNILEVER NV.
 PA (UNIL) UNILEVER PLC.
 PI Kils FM, Schreuder MP, Toschka H, Verrips CT;
 DR WPI: 1994-035071/04.
 DR N-PSDB; AA054012.
 XX
 PT Immobilisation of enzymes to microbial cell wall - by prodn. of
 PT fusion protein of enzyme linked to anchoring protein
 XX
 PS Example 1; Page 32-39; 99pp; English.
 XX
 CC The alpha-agglutinin is used in a method to immobilise enzymes to a
 CC microbial cell wall. The coding sequence is used in the production
 CC of a recombinant polynucleotide which comprises a structural gene
 CC encoding a protein with catalytic activity and at least part of a
 CC gene encoding at least the C-terminus of a protein capable of
 CC anchoring in a eukaryotic or prokaryotic cell wall. The anchoring
 CC fragment or protein is selected from alpha agglutinin, AGA 1, FLO 1,
 CC major cell wall protein of lower eukaryotes or a proteinase of
 CC lactic acid bacteria. The recombinant polynucleotide preferably
 CC also comprises a sequence encoding a signal peptide to ensure
 CC secretion of the expressed product. The signal peptide is
 CC preferably derived from glycosyl-phosphatidylinositol, anchoring
 CC protein, alpha factor, alpha-agglutinin, invertase or inulinase,
 CC alpha-amylase of *Bacillus* or proteinases of lactic acid bacteria.
 CC The host microorganism can be used for performing enzymatic
 CC processes on an industrial scale.
 XX
 SQ Sequence 650 AA;

Query Match 6.0%; Score 114; DB 15; Length 650;
 Best Local Similarity 20.9%; Pred. No. 0.087;
 Matches 98; Conservative 52; Mismatches 147; Indels 172; Gaps 23;

QY 4 ILFTFTLFSSVL---ETFA-----VSADKIPG-----DESTINIFGPRDRNE-- 43
 DB 8 llwflslalaaaindittfletptankpqdgvtatdfdaasslregdefl 67
 QY 44 SSP---KHNIINNHITA-----YSSSHLYDRMFLC-----LSHNTLN 80
 DB 68 smphvyrtiklinsqatalsadgleafkcyvsqgaaylyentftctcaqnllssynld 127
 QY 81 GACPTSENPSSSVSGETNTTLQFTEKRLIKRELQIKGYKQLLFKSVNC-PSGLTNSLA 139

D 128 gsfslsfdsdgssayeyel-----enakkfksgpmlvklngmsdvvnfdpaftenvf 182
Q 140 HFNCKNNAASASLYL--YIPAGELKNLPFGIMDATLTKLRKRRYSEYGYRINIRIK 197
D 183 hsgstgysfesyhlgnycpnyf----lsg-----teki-----dydsannvnl- 225
Q 198 LTDGNIOI-----WLPQFKSDARVDLNLRPFGGTYIGRN---SYDMCFYDQ--- 242
D 226 --dcssvgyvssndfndwfpqsyndlnady-----tcfgslwlitldekltydgeml 275
Q 243 -----YSTNSSLE--IRFGDNN--PKSDCKF 265
D 276 wvnalgsiparvntldhalefgyrcldtianttyatqfstreflyqgnlgrtasaks 335
Q 266 YLRKINDTKEI---AVTIS-----LLAGKSLPPTNGSTLINADAA 304
D 336 fisltdltslntsaystgststveigtntsevisahvtscklaptatsltiaqts 395
Q 305 SLENNMN-----RITAVTMPEISVPLCWPGRL 332
D 396 lyscdsnltvgydltltsevisdvtelstrelastvaapsttqwtgam 444

RESULT 5

AAB01833: standard; Protein; 1005 AA.

AAB01833:

11-SEP-2000 (first entry)

Haemophilus influenzae strain K21 mature HMW2A protein, SEQ ID NO:41.

Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight; non-typeable Haemophilus influenzae; NTHI; non-encapsulated;

human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis; detection; diagnosis.

Haemophilus influenzae strain K21.

MO200020609-A2.

13-APR-2000.

07-OCT-1999; 99WO-CA00938.

07-OCT-1998; 98US-0167568.

08-DEC-1998; 98US-0206942.

(CONN-) CONNACHT LAB LTD.

Loosmore SM, Yang Y, Klein MH;

WPI: 2000-303789/26.

N-PDB; AAA52182.

Nucleic acid molecule for producing recombinant high molecular weight proteins of Haemophilus which are used as a vaccine to provide protection against Haemophilus induced diseases in humans -

Claim 8: Fig 21A-C: 307pp: English.

The invention relates to the recombinant production of Haemophilus influenzae high molecular weight (HMW) proteins in Escherichia coli. The expression construct used to effect recombinant expression comprises a promoter functional in E. coli (e.g., the T7 promoter) operably linked to a modified hmwABC operon from a non-typeable (non-encapsulated) H. influenzae (NTHI). Most HMW-expressing NTHI strains contain two hmw gene clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA, hmwB and hmwC genes. The hmwA genes encode the structural HMW proteins and the hmwB and hmwC genes encode accessory proteins which are responsible for post-translational processing and secretion of the HMW

CC proteins. The modified hmwABC operon used in the expression construct of the invention contains an A gene modified such that it encodes only the mature HMW. The invention also discloses hmwA genes (AAA52175-A52198) and HMW proteins (AAB01832-A01849) from the non-typeable H. influenzae strains Joyce, K1, K21, UCC02, PMH1, 15 and 12. The nucleic acids and vectors are used for the production of recombinant H. influenzae HMW proteins which can be used as vaccines to mediate a humoral or cell-mediated immune response to provide protection against diseases in humans caused by H. influenzae (e.g., otitis media, epiglottitis, pneumonia and tracheobronchitis). The HMW proteins are also useful as antigens in immunoassays for detecting antibodies against Haemophilus, HMW proteins and/or HMW peptides. The nucleotide sequences encoding the HMW proteins can be used to isolate and clone hmw genes from other non-typeable strains of Haemophilus via hybridisation reactions. The present sequence represents a mature HMW protein from a non-typeable strain of H. influenzae.

Sequence 1005 AA;

Query Match

Best Local Similarity 20.3%; Pred. No 0.28; Length 1005; Matches 68; Conservative 65; Mismatches 147; Indels 55; Gaps 14.

Q 60 ESHTLYDRMTFLCSHNTLWACPTSENPSSVSGETNTTQTEKRS-LIKRELQIK 118

D 363 eknaifsthnltllygnvltlgenssnikgnlnlnskanvltqahagtsltdk----- 418

Q 119 GYKOLLFRKSVNCPGSLTL-NSAHFNCKNNAASASLYLIPAGEL-KNLPFGIWL-DA 173

D 419 --rtcltignvsvgnlnlignahldgnlslaesakf-----gqktnmnlitqftnng 471

Q 174 TLKLRVKRRSEYGTGTINTIKLTD-----KGNL-----QIWLQFQSDARVD 218

D 472 tadnltggyvklqgdltmngnltnasvngqtltnnglnltnkkgdlnkdklkaeig 531

Q 219 L-NLRPTGGGTYIGRNSVDMCFYDGYSTNSSSLEIRFQDNNPKSD-GKEYLRKINDTK 275

D 532 lgnlsgkegnlttsdklnl-----tkrlelkadtdqgsdgyasnanliltkt 582

Q 276 EIAATLSLLAG---KSLTPTNGSTLINADAASETMWNITAVTMPEISVPLCWPGRL 332

D 583 eltltdlnlsgfknkaeltakdsdlilgkassdmsnakqitldkvdksks--agnhvv 640

Q 333 QLDKAVENPEA-----GOYGNINVTFTPSSQTL 361

D 641 tlmskvtlsgdstgtsngsdmnlgtltsakdvtrv 675

RESULT 6

AAB01832: standard; Protein; 1011 AA.

AAB01832:

11-SEP-2000 (first entry)

Haemophilus influenzae strain K21 HMW2A protein, SEQ ID NO:39.

HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight; non-typeable Haemophilus influenzae; NTHI; non-encapsulated;

recombinant production; Escherichia coli; antibacterial; vaccine; human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis; detection; diagnosis.

Haemophilus influenzae strain K21.

MO200020609-A2.

13-APR-2000.

07-OCT-1999; 99WO-CA00938.

DB 1002 nadakk---vtfcdkvkxskistdghnvtlnsevktsngssnagnndnst 1046

RESULT 8

AAW30292 ID AAW30292 standard; Protein; 1601 AA.

AAW30292;

14-APR-1998 (first entry)

Non-typeable Haemophilus high mol.wt. surface protein HMW4.

Non-typeable Haemophilus; high molecular weight surface protein; HMW4; immunogen; vaccine; otitis media.

Haemophilus influenzae strain 5.

Key Location/Qualifiers

Misc-difference 372 /note= "encoded by TCT"

Misc-difference 400 /note= "encoded by AAT"

WO9736914-A1.

09-OCT-1997.

01-APR-1997; 97WO-US04707.

01-APR-1996; 9605-0617697.

(BARE/) BARENKAMP S J.

Barenkamp SJ;

WPI; 1997-503038/46.

N-PSDB; AAT909933.

High molecular weight proteins of non-typeable Haemophilus influenzae - useful for vaccine production

Claim 1; Page 97-102; 183pp; English.

This protein comprises the high molecular weight surface protein HMW4 (123 kDa) of non-typeable Haemophilus influenzae strain 5 that has the immunological ability to protect against disease caused by a non-typeable Haemophilus strain and is characterised by at least one surface-exposed B-cell epitope that is recognised by monoclonal antibody AD6. The HMW4 amino acid sequence was deduced from an isolated hmw4 gene (see AAT909933). HMW1 (see AAW30293), HMW2 (see AAW30294) and HMW3 (see AAW30291) have also been identified. A conjugate comprising HMW4 linked to an antigen, hapten or polysaccharide, and a synthetic peptide of 6-150 amino acids corresponding to at least protective epitope of HMW4 are also claimed. HMW proteins, conjugates and peptides can be used in vaccines, as immunogens for preparation of antibodies and as antigens for detection of these antibodies.

Sequence 1601 AA;

Query Match 5.9%; Score 110.5; DB 18; Length 1601;

Best Local Similarity 20.7%; Pred. No. 0.71; Mismatches 127; Indels 87; Gaps 18;

Matches 72; Conservative 62; Mismatches 127; Indels 87; Gaps 18;

48 ANILNNHTAYSESHTLYRMFMFLCSNNTNGACPTSENPSS-----SVSGETNITL 102

816 yneyskh--alnshnl-----tlgnvntlg-----enssiltgnlntknkavtl 862

103 QF-----TEKRSLLKELOIKYKQLFKSVNCPGSLTTLNSAHFNCKNAASASLY 154

863 gadtssnntglkrtltlgnisvegnlsitgnanivnlsf--aedstfkgsa----- 915

155 LVIAPGELKNLPRGIIW--DATLKRKRRESEPTGYT-----INTIKLT----- 199

916 -----dnltitgltfmgntaninkgvykvgldinkgldntltasgtqkling 967

200 ---DKGNIOIWLPOKSDARVDL--NLRPPTGGGYIGNSVDMCPYDSTNSSLLEIR 253

968 nltneggdlnl--knlkadaeiglgisqkegnltsdskvnl-----lntgltkeg 1018

254 FQ----DNNPKSDGKRYLKRINDTKEIAYTSLLAG---KSLPTNGTSLNIADASL 306

1019 veggirsdseseaenaltlg-----tkelklagldlnisgfkaelitkngsdltlignasg 1073

307 EFTMNRITAVTMEPISVPVICMPG--RLQDPAKVENPEAGQVGNINVT 353

1074 nadakk---vtfcdkvkxskistdghnvtlnsevktsngssnagnndnst 1118

RESULT 9

AA15945 ID AA15945 standard; Protein; 2383 AA.

AA15945;

05-OCT-2000 (first entry)

E. coli proliferation associated protein sequence SEQ ID NO:302.

Escherichia coli; E. coli; proliferation; inhibition; screening; antimicrobial; bacterial growth; antisense therapy; antibacterial.

Escherichia coli.

WO200044906-A2.

03-AUG-2000.

27-JAN-2000; 2000WO-US02200.

27-JAN-1999; 99US-0117405.

(ELIT-) ELITRA PHARM INC.

Zyskind J, Ohlsen KL, Trowick J, Forsyth RA, Froelich JM, Carr GJ;

Yamamoto RT, Xu HH;

WPI; 2000-514822/46.

N-PSDB; AAA65950.

Novel polynucleotides and polypeptides associated with microorganism proliferation, used to identify inhibitors of bacterial growth and proliferation, for use in antisense therapy -

Claim 11; Page 224-229; 316pp; English.

AAA5809 to AAA6589 and AAA66058 to AAA66138 represent nucleotide sequences derived from Escherichia coli which inhibit E. coli proliferation. AAA6589 to AAA66055 and AA15886 to AA16040 represent nucleotide and protein sequences associated with E. coli proliferation. CC AAA66056 and AAA66057 represent primers used for sequencing E. coli proliferation inhibiting nucleotide inserts in an example from the CC present invention. Methods from the present invention can be used to identify a proliferation-required gene in a microorganism, by contacting a microorganism with a proliferation-required gene activity inhibitory CC nucleic acid identified in another organism, and determining if CC inhibition occurs in the second microorganism. The nucleic acid sequences identified as being required for bacterial growth and proliferation, can be used for antisense therapy for killing bacteria.

Sequence 2383 AA;

Query Match 5.8%; Score 109; DB 21; Length 2383;

RESULT	10
AAB01835	
ID	AAB01835 standard; Protein; 1095 AA.
AC	
AA01835;	
DT	11-SEP-2000 (first entry)
DE	Haemophilus influenzae strain LCDC2 mature HMW1A protein, SEQ ID NO:45.
XX	
XX	Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW	non-tyeable Haemophilus influenzae; NTHi; non-encapsulated;
KW	recombinant production; Escherichia coli; antibacterial; vaccine;
KW	human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW	detection; diagnosis.
XX	
OS	Haemophilus influenzae strain LCDC2.
XX	
PN	WO200020609-A2.
XX	
PD	13-APR-2000.
XX	
PF	07-OCT-1999; 99WO-CA00938.
XX	
PR	07-OCT-1998; 98US-0167568.
PR	08-DEC-1998; 98US-0206942.
XX	
PA	(CONN-) CONNAUGHT LAB LTD.
XX	
PI	Loosmore SM, Yang Y, Klein MH;
XX	
DR	WPI: 2000-303789/26.
DR	N-PSDB: AAA52184.
XX	
PT	Nucleic acid molecule for producing recombinant high molecular weight
PT	proteins of Haemophilus which are used as a vaccine to provide
PT	protection against Haemophilus induced diseases in humans -
XX	
PS	Claim 8; Fig 22A-P; 307pp; English.
CC	The invention relates to the recombinant production of Haemophilus

```

Query Match Summary              5.7%; Score 107; DB 21, Length 1095;
Blast Local Similarity 24.6%; Pred No. 0.86;
Matches 71; Conservative 44; Mismatches 110; Indels 64; Gaps 15.

OY 45 SFRKHLNNHTAYAESHTLYD-----RMTFLCLSSHNTLNGACPTSENPSSSS 93
      | | | | | : | | | | : | | | : | | | : | | | : | | |
DB 281 sklpplva---vhdgnlthlfnngvsvlgggdvnfhnsssnhvcbylvsqfnaae 337
      | | | | | | | | | | | | | | | | | | | | | | | | |

OY 94 VSGEINTLTQFEKRSKLKRELQIKGVKQLKRSVNCPSGLTNSAHFVCKNNAASGASL 153
      | | | | | | | | | | | | | | | | | | | | | | | | |
DB 338 gs-----slrf-----ksgstrfal---tiesdlfnatgaislnyvgaidg 378
      | | | | | | | | | | | | | | | | | | | | | | | | |

OY 154 YLYIPAGELKLNPFSGIMDATLKLKVRKRRYSFTGYTYT---NTTKLTDRGNIQIWP 209
      | | | | | | | | | | | | | | | | | | | | | | | | |
DB 379 nlqslvanaknltfeg---gnltlaadkklpeltkgnltvkeganvltlrseanyndksal- 434
      | | | | | | | | | | | | | | | | | | | | | | | | |

OY 210 QEKSPARVDLNRPTGGGYIGRN-----SYDMCYDYDGYSTSSSLSEIRFOONNRKD-- 262
      | | | | | | | | | | | | | | | | | | | | | | | | |
DB 435 slrgvntkngnltvtgsalnleknltlvegsakflapnysflvsvgl-----fdnqgksms 490
      | | | | | | | | | | | | | | | | | | | | | | | | |

OY 263 --GKFLRKINDDTKEIAVYTLs-----LLLAGSKLPTPTNGTSLNIAD 302
      | | | | | | | | | | | | | | | | | | | | | | | | |
DB 491 lakgahfekdln-nlkslnlttssdaytlleg-nlttng-dnltld 536
      | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 11
ID AAB01834 standard; Protein; 1101 AA.
XX AAB01834:
XX
XX AAB01834:
XX
XX 11-SEP-2000 (first entry)
XX
DE Haemophilus influenzae strain LCDC2 HMW1A protein, SEQ ID NO:43.
XX
XX
XX HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
XX non-typable Haemophilus influenzae; NTH1; non-encapsulated;
XX recombinant production; Escherichia coli; antibacterial; vaccine;
XX human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
XX detection; diagnosis.
XX
XX Haemophilus influenzae strain LCDC2.
XX
XX WO200020609-A2.
XX
XX

```

PD 13-APR-2000.
 XX
 PF 07-OCT-1999; 99WO-CA00938.
 XX
 PR 07-OCT-1998; 98US-0167568.
 PR 08-DEC-1998; 98US-0206942.
 XX
 PA (CONN-) CONNAUGHT LAB LTD.
 XX
 PI Loosmore SM, Yang Y, Klein MH;
 XX
 DR WPI: 2000-303789/26.
 DR N-PSDB; AAA52183.
 XX
 PT Nucleic acid molecule for producing recombinant high molecular weight
 PT proteins of Haemophilus which are used as a vaccine to provide
 PT protection against Haemophilus induced diseases in humans -
 XX
 PS Claim 12; Fig 22A-P; 307pp; English.
 XX
 CC The invention relates to the recombinant production of Haemophilus
 CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The
 CC expression construct used to effect recombinant expression comprises a
 CC promoter functional in E. coli (e.g., the T7 promoter) operably linked
 CC to a modified hmwaBC operon from a non-typable (non-encapsulated) H.
 CC influenzae (NTH1). Most HMW-expressing NTH1 strains contain two hmw gene
 CC clusters termed hmwaABC and hmw2ABC. Each hmwaBC operon comprises hmwa,
 CC hmwB and hmwC genes. The hmwa genes encode the structural HMWA proteins
 CC and the hmwB and hmwC genes encode accessory proteins which are
 CC responsible for post-translational processing and secretion of the HMWA
 CC proteins. The modified hmwaBC operon used in the expression construct of
 CC the invention contains an A gene modified such that it encodes only the
 CC mature HMWA. The invention also discloses hmwa genes (AAA52175-A52198)
 CC and HMWA proteins (AAB01824-B01849) from the non-typable H. influenzae
 CC strains Jovc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
 CC vectors are used for the production of recombinant H. influenzae HMW
 CC proteins which can be used as vaccines to mediate a humoral or
 CC cell-mediated immune response to provide protection against diseases in
 CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,
 CC pneumonia and tracheobronchitis). The HMW proteins are also useful as
 CC antigens in immunoassays for detecting antibodies against Haemophilus,
 CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
 CC HMW proteins can be used to isolate and clone hmw genes from other
 CC non-typable strains of Haemophilus via hybridisation reactions. The
 CC present sequence represents an HMWA protein from a non-typable strain of
 CC H. influenzae.
 XX
 SO Sequence 1101 AA;

DB	497	jakgahnfndin-nkcslnitnssayrtllieg-nitnsg-dlnitd	542
	RESULT	12	
	AAR85290		
ID	AAR85290	standard; Protein; 671 AA.	
XX			
AC	AAR85290;		
XX			
DT	02-APR-1996	(first entry)	
XX			
DE	Streptococcus faecalis autolysin.		
XX			
XX	Lysin; autolysis; culture; lactic acid bacteria; fermentation;		
KW	cheese; foodstuffs; induction.		
XX			
OS	Streptococcus faecalis.		
XX			
PN	W09531561-A1.		
XX			
PD	23-NOV-1995.		
XX			
PF	12-MAY-1995;	95WO-NL00170.	
XX			
PR	12-MAY-1994;	94EP-0201353.	
XX			
PA	(UNIL) QUEST INT BV.		
XX			
PI	Buist G, Kok J, Ledebor AM, Venema G;		
XX			
DR	WPI: 1996-010946/01.		
XX			
PT	Lysis of a culture of lactic acid bacteria in, e.g. cheese		
PT	production - by in situ prodn. of an auto:lysin, regulated by an		
PT	inducible promoter.		
XX			
PS	Disclosure: Page 66-69; 103pp; English.		
XX			
CC	In situ production of a homologous autolysin or a heterologous		
CC	autolysin from a food grade Gram positive bacteria, can be used in		
CC	a process for the lysis of a culture of lactic acid bacteria. The		
CC	process can be used in the manufacture of products containing		
CC	cultures of lactic acid bacteria e.g. cheese, where the culture is		
CC	lysed following the completion of fermentation. The enhanced		
CC	induction of the autolysin is performed some hours after the		
CC	fermentation is finished. No extra lysin needs to be added and the		
CC	lysin does not need to be isolated or encapsulated. The time of		
CC	lysin can be precisely controlled. This is the Streptococcus		
CC	faecalis autolysin.		
XX			
SO	Sequence	671 AA;	

```

Query Match Similarity      5.7%: Score 107; DB 21; Length 1101;
Best Local Similarity      24.6%: Pred No. 0.87;
Matches 71; Conservative 44; Mismatches 110; Indels 64; Gaps 15

QY      45  SPKANHNNHITAYAESHTLYD-----RMFFLLCSHNTLNGCAPTSENPSSS 93
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      287  slkxplvsn-----vhdgnhltfngnvsylvggdvntflnsssnhwhchgyvlskqntnase 343
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      94  VSGEINIIQTFEKSRLKRELQIKGYKOLLKFSVNCPSGGLTNSAHFNCKNKAASGL 153
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      344  gs-----slrf-----ksegstrtaf-----flesdltlnacgmslnqvagld 384
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      154  YIYIPAGELKNLPFGGIMPATLKLARKRKRYSEIYTYT-----NITIKLTDKGNIQIIMP 209
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      385  nqkxlvanknltfeg-----gnltlaadkxplkignltvkeganvllrsanvgndksal- 440
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      210  QFKSDARYDLNITPRGGGTGYIGRN-----SYDWCIFYDGYSTNSSLPIRQDNNPKSD-- 262
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      441  sirgnvtknqnlvtvrgsainieknltvgsakliannpysfnvsqf-----fdagqksnis 496
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      263  ---GKFIYLRKINDDTKEIAYTLS-----LLAGKSLIPTNGSTSLNIAD 302
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Query Match          5.6%  Score 106;  DB 17;  Length 671;
Best Local Similarity 23.6%  Pred. No. 0.51;
Matches 73;  Conservative 39;  Mismatches 103;  Indels 94;  Gaps 17.

QY      41 RMSSPKHHI-LNNHITASESHIYIDRTEFLCLSHNLTNACPTSENPSSSV----- 94
      | : : : | | | | | | : : : : | : | : : : | : : :
Db      309 ryatdpnyaklnnvltay--nltyddcpsggnitgvgvtpgctgsmgsgntnytyvk 366
      | : : : | : : : | : : : | : : : | : : : | : : :
QY      95 SGET--NITLQF-----TERSL--IKRELQIKGYKQLFKSYNCSGLTLNLAHFENCKNN 146
      | : : : | : : : | : : : | : : : | : : : | : : :
Db      367 sgdtlnklaaagygsवानलरswngisgdlifvgqklivkga-----sgntgsgsgnsgnnn 423
      | : : : | : : : | : : : | : : : | : : : | : : :
QY      147 AASGASLYLIPAGELKNLPFGIGWDATLKLVRKRRYSSETGYTYTINITIKLTDKGNIOI 206
      | : : : | : : : | : : : | : : : | : : : | : : :
Db      424 -qsgntyytvmksgdln-----klaagy-----vlt-----anlr 455
      | : : : | : : : | : : : | : : : | : : : | : : :
QY      207 WLPOKSDARVDLNL-----RPTGGGTYIGRNSVDMCFDGYSTNSSLFIRQDNN 258
      | : : : | : : : | : : : | : : : | : : : | : : :
Db      456 w-----ngisgdlifvgqklivkkgtsyut-----gsgnsgs-----nnn 491

```

QY	259	PKSGKFLRKINDTKEA--TTLIS-----LLAGKSLPTFNGSLINAA	303
Db	492	qsgtncrytcllsgdltlnktaagyvsvaallrswngisgdlftagkxlivkktsntggs	551
QY	304	ASLETNNMR 312	
Db	552	sngsgnmng 560	
RESULT 13			
AAAB01830			
XX	AAAB01830	standard; Protein; 1222 AA.	
XX	AC		
XX	AD		
XX	DT		
DE	11-SEP-2000	(first entry)	
XX			
KW	H. influenzae strain K1 mature full-length HMW1A protein, SEQ ID NO:37.		
KW	Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;		
KW	non-lysable Haemophilus influenzae; NTHi; non-encapsulated;		
KW	recombinant production; Escherichia coli; antibacterial; vaccine;		
KW	human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;		
KW	detection; diagnosis.		
XX			
OS	Haemophilus influenzae strain K1.		
XX			
XX	Key	Location/Qualifiers	
FT	Misc-difference	307	
FT		/note- "Encoded by GG"	
XX	MO200020609-A2.		
XX	13-APR-2000.		
PE	07-OCT-1999;	99MO-CA00938.	
XX			
PR	07-OCT-1998;	98US-0167568.	
PR	08-DEC-1998;	98US-0206942.	
XX			
PA	(CONN-) CONNAUGHT LAB LTD.		
XX			
PI	Loosmore SM, Yang Y, Klein MH;		
DR	WPI: 2000-303789/26.		
DR	N-PSDB: AAA52180.		
XX			
XX	Claim 8; Fig 20A-R; 307pp; English.		
XX			
XX	The invention relates to the recombinant production of Haemophilus		
XX	influenzae high molecular weight (HMW) proteins in Escherichia coli. The		
XX	expression construct used to effect recombinant expression comprises a		
XX	promoter functional in E. coli (e.g., the T7 promoter) operably linked		
XX	to a modified hmwABC operon from a non-typable (non-encapsulated) H.		
XX	influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene		
XX	clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,		
XX	hmwB and hmwC genes. The hmwA genes encode the structural HMW proteins		
XX	and the hmwB and hmwC genes encode accessory proteins which are		
XX	responsible for post-translational processing and secretion of the HMW		
XX	proteins. The modified hmwABC operon used in the expression construct of		
XX	the invention contains an A gene modified such that it encodes only the		
XX	mature HMW. The invention also discloses hmwA genes (AAA52175-A52198)		
XX	and HMW proteins (AAAB01824-B01849) from the non-typable H. influenzae		
XX	strains Joyce, K1, K21, LCD2, PMH1, 15 and 12. The nucleic acids and		
XX	vectors are used for the production of recombinant H. influenzae HMW		
XX	proteins which can be used as vaccines to mediate a humoral or		
XX	cell-mediated immune response to provide protection against diseases in		
XX	humans caused by H. influenzae (e.g., otitis media, epiglottitis,		
XX	pneumonia and tracheobronchitis). The HMW proteins are also useful as		

CC	antigens in immunoassays for detecting antibodies against Haemophilus,
CC	HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC	HMW proteins can be used to isolate and clone hmw genes from other
CC	non-typable strains of Haemophilus via hybridisation reactions. The
CC	present sequence represents a mature HMW protein from a non-typable
CC	strain of H. influenzae.
XX	
SO	Sequence 1222 AA;
	Query Match 5.6%; Score 106; DB 21; Length 1222;
	Best Local Similarity 20.2%; Pred. No. 1,3; Mismatches 165; Indels 128; Gaps 21
Matches	91; Conservative 67; Mismatches 165; Indels 128; Gaps 21
OY	11 PFSSVLFPTFAVSAD-----KIPGDESTINIFGRDRNESSPKHNI-----LNNHTAY 58
Dd	265 fndtvlfaaaasavnnfklppldvktltn-----ghclfkgnslvggadvnfhnfs 318
OY	59 SESHTLYDPMFTLCSSHN-----TLNCA 82
Dd	319 ssnygtlyg----vllsqnffsaagssalkfksegsthaaflikndlllnatgnslngy 374
OY	83 CTSNPSPSSSSSGEN-----NTITQFPEKSLIKREIQINGYKQLPRSVNC----- 130
Dd	375 agidsnlkkslianknitfgsnltlaackkpelkgnlvtkeganavltsanyndksa 434
OY	131 -----PSGLTINSAHFNCNK-N-AASGASLYLTYPAGELNLPF--GGIWDATLKLR 178
Dd	435 lsirgvnckngiltvtgasinlekhltvegsakflap-----nysfnvsjlfddgskn 489
OY	179 VK-----RRYSFTGYTYTINTITIKLTD-----KGNIQTWLDPFKSDARVDNLRLPFGG 227
Dd	490 lsiakgaalfkdientlsgnsitlctsdshhtlkgnl-----tnrkgdnlntngdn 541
OY	228 T-YYGRNSVDMCFDGVSTNSS-----LEIRQDNPKSD-GKFYLKRINDOTKEI 277
Dd	542 teiqdggn---lsqkgnltsdsdkvnleritlkagvngdsdneatsanlttkel 598
OY	278 AYTLLLAG--KSILTPNGTSLNIADAASL-ETNMNRITAVVTMPDISVPVLCPGRIQ 333
Dd	599 kltndlninsgfkaeltakdsnltlctgdsdgnctadakkvtsnvvdskisas--dhnyt 656
OY	334 LDAKVE--NPEAGQMGYNINVTFPSSQTL 361
Dd	657 lnskvetsgdstedggnmtgltitaknv 687
RESULT 14	
AABO1828	
ID	AABO1828 standard; Protein: 1228 AA.
AC	AABO1828;
DT	11-SEP-2000 (first entry)
DE	Haemophilus influenzae strain K1 full-length HMW1A protein, SEQ ID NO:34.
XX	
KW	HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW	non-typable Haemophilus influenzae; NTHi; non-encapsulated;
KW	recombinant production; Escherichia coli; antibacterial; vaccine;
KW	human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
CC	detection; diagnosis.
XX	
OS	Haemophilus influenzae strain K1.
FH	Key Location/Qualifiers
FT	Misc-difference 313 /note= "Encoded by GC"
XX	
FN	WO200020609-A2.
DD	13-APR-2000.

PF 07-OCT-1999; 99WO-CA00938.
 PR 07-OCT-1998; 98US-0167568.
 PR 08-DEC-1998; 98US-0206942.
 XX
 FA (CONN-) CONNAUGHT LAB LTD.
 XX
 PI Loosmore SM, Yang Y, Klein MH;
 XX
 DR WPI: 2000-303789/26.
 DR N-PSDB; AAA52179.
 XX
 PT Nucleic acid molecule for producing recombinant high molecular weight
 PT proteins of Haemophilus which are used as a vaccine to provide
 PT protection against Haemophilus induced diseases in humans -
 XX
 PS Claim 12; Fig 20A-R; 307pp; English.

CC The invention relates to the recombinant production of Haemophilus
 CC Influenzae high molecular weight (HMW) proteins in Escherichia coli. The
 CC expression construct used to effect recombinant expression comprises a
 CC promoter functional in E. coli (e.g., the T7 promoter) operably linked
 CC to a modified hmwaB operon from a non-typeable (non-encapsulated) H.
 CC Influenzae (NTHI). Most HMW-expressing NTHI strains contain two hmw gene
 CC clusters termed hmw1ABC and hmw2ABC. Each hmwaB operon comprises hmwa,
 CC hmwb and hmwc genes. The hmwa genes encode the structural HMWA proteins
 CC and the hmwb and hmwc genes encode accessory proteins which are
 CC responsible for post-translational processing and secretion of the HMWA
 CC proteins. The modified hmwaB operon used in the expression construct of
 CC the invention contains an A gene modified such that it encodes only the
 CC mature HMWA. The invention also discloses hmwa genes (AAA52175-A52198)
 CC and HMWA proteins (AAB01824-B01849) from the non-typeable H. Influenzae
 CC strains Joyce, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
 CC vectors are used for the production of recombinant H. Influenzae HMW
 CC proteins which can be used as vaccines to mediate a humoral or
 CC cell-mediated immune response to provide protection against diseases in
 CC humans caused by H. Influenzae (e.g., otitis media, epiglottitis,
 CC pneumonia and tracheobronchitis). The HMW proteins are also useful as
 CC antigens in immunoassays for detecting antibodies against Haemophilus,
 CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
 CC non-typeable strains of Haemophilus via hybridisation reactions. The
 CC present sequence represents an HMWA protein from a non-typeable strain of
 CC H. Influenzae.

CC Sequence 1228 AA;

Query Match 5.6%; Score 106; DB 21; Length 1228;
 Best Local Similarity 20.2%; Pred. No. 1.3;
 Matches 91; Conservative 67; Mismatches 165; Indels 128; Gaps 21;

OY 11 FFSVLTFFFAVSAD-----KIPGDESITNIFGRDRNESSPKHNI-----LNHITAY 58
 DB 271 fndvtvtnhaasavafnkplvdkvtn-----gnhtlfgnsvlvggdvtnhfnas 324
 OY 59 SESHTLYDRMTFLCLSHN-----TLNGA 82
 DB 325 ssnlytly-----vllsegnfaasgsslkfkfseghaafctikndllinatgnslngv 380
 OY 83 CPTSENPSSSSYSGET-----NITLOFTKRSILKRELQIKGYQLFKFSYVC----- 130
 DB 381 agidsnllksllaanfkfegntllaadkpklekgntlvkeganvclrsanngdksa 440
 OY 131 -----PSGLTNSAHFNCNKN-AASGASLYLYPAGELKNPF--GGIWDATLKLIR 178
 DB 441 lsiqgnvtnkgnltvcsaineknlvvesakflamp-----nysltnvsqglfdngqsgn 495
 OY 179 VK-----RRYSEYTGTYINITIKLTD-----KGNIOIWLQFQSDARVDNLPTGGG 227
 DB 496 lsiakgaifkfdientglnlttkdsnhltikgnl-----tnkgdlnitnngdn 547
 OY 228 T--YIGRNSVDMCFYDGYSTNSS-----LEIRFQDNPKSD-GKFFLRKINDDTKEI 277

DB 548 teiqign-----isqkegnlhtssdkvnterltlkagvngdsdsneatsanltktkel 604
 OY 278 AVTSLSLANG---KSLFTPTNGTSLNIADAASL-ENNMRTIAVAMPETISVYVLCMPGRLO 333
 DB 605 klndlnisgfnkaeltakdsnltltdgnsdagndakkvlfsvnkdkiskas--dhnvt 662
 OY 334 LDAKVE---NPEAGQYMGNIWVTFPPSSQTL 361
 DB 663 lnskvetsgdtdstedggnmtgtltaknv 693

RESULT 15
 AAR41731
 ID AAR41731 standard; Protein; 1338 AA.

AC AAR41731;
 DT 26-APR-1994 (first entry)
 XX
 DE High molecular weight protein 3 (HMW3).

KW HMW; high molecular weight protein; virus; vaccine; influenza;
 KW epitope; immunity; haemophilus influenzae.

OS Haemophilus influenzae.

PN W09319090-A.

XX 30-SEP-1993.

PD 16-MAR-1993; 93WO-US02166.

PF 16-MAR-1992; 92GB-0005704.

XX 16-MAR-1992; 92GB-0005704.

PR (BARE/) BARENKAMP S J.
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX Barenkamp SJ;
 PI WPI: 1993-320683/40.

DR N-PSDB; AAQ49510.

XX High molecular weight surface proteins - of non-typeable
 XX haemophilus which exhibit immunogenic properties

PT Claim 5; Figure 10; 100pp; English.

CC The isolation and purification of the high molecular weight protein
 CC enables the identification of the major protective epitopes of the
 CC protein by conventional epitope mapping. These epitopes can then be
 CC synthesised using standard techniques and incorporated into fully
 CC synthetic or recombinant vaccines.

XX Sequence 1338 AA;

Query Match 5.4%; Score 101; DB 14; Length 1338;
 Best Local Similarity 21.5%; Pred. No. 4.3;
 Matches 90; Conservative 52; Mismatches 172; Indels 104; Gaps 18;

OY 21 VSADKIPDESITNIFGRDRNESSPKHNI-LNHI-TAYSESHLYDRMTFLCLSHN----- 74
 DB 817 vtfdkv-----kdekstidgtnvlnsevktsngsnagnsdngstgltsakdv 864
 OY 75 -----SHNTLN-----GACPTSENPSSSSYSGETNTILO-----FTEKR 108
 DB 865 tvnnvnvtshklnlnsaagntlkgtlnatgsvetraqnltkgnltqnvtylate 924
 OY 109 SLIKRELQI---KGYQLFKFSYVCPGLTNSAHFNCNKNNAASGASL----- 153
 DB 925 nlvttenavlnatsgtvnlsctkgdtkgltsgtsgnvnlt---tasgnltkvsnltgqvt 981

```

OY 154 -----LYIPAGELKNLPFGIMDATLKRVRRYSEYGYTYTINITYIKLTDKNGIQIW 207
      | | : | : | : | : | : |
Db 982 vtadagalttagstlsatgnaitktgdingkveassgvltvatgatlavnglsgn 1041
      | | : | : | : | : | : |
OY 208 LPQKSDARVDLNLRPPTGGGYIGRNSVDMCFYDGYSTNSSLLEIRFODNN--PKSDGK 264
      : | : | : | : | : | : |
Db 1042 tvltadsg--kltstvgstlngtnsvt-----ssqsgdieglisgnltvnvtastgd 1092
      : | : | : | : | : | : |
OY 265 FYL-RKINDTKEIATLS-----LLLAGKSITPTNGTSLNITADAASLETNN--RITA 315
      : | : | : | : | : | : |
Db 1093 ltignsakveakngaatllaesgkltqgssltssngqtlitakdsiagninaanvtl 1152
      : | : | : | : | : | : |
OY 316 VFMPEISV---PVLGWPGRLOL--DAKVENPEAG-----QXGNINVTFTPPSS 358
      : | : | : | : | : | : |
Db 1153 ntgtlilttgdkinatsgtlitinakdaxidgaasgdlrtvvnatnasgsgnvtaktss 1210

```

Search completed: July 1, 2002, 11:24:10
 Job time: 317 sec

This Page Blank (uspto)

[illegible]

Db 1 MNKILFTLFFSSGFTTFAVADKNPGENMTITQPHDRGSSPIYINILNSYLAAYNG 60
QY 61 SHTLYDMRTFLCLSSHTNLNGACPTSENPSSSVSGETNITLQFTERSLIKRELOIKGY 120
Db 61 SHHLYDMRTFLCLSSHTNLNGACPTSENPSSSVSGETNITLQFTERSLIKRELOIKGY 120
QY 121 KOLLFKNVNPSSQTLNLSAHPNCKNAASGLYLYTPAGELKNLPFGGIMDATLKIRVK 180
Db 121 KOLLFKNVNPSSQTLNLSAHPNCKNAASGLYLYTPAGELKNLPFGGIMDATLKIRVK 180
QY 181 RRTSETGTGTINTITIKLTQKNGIQIOLPOFKSDARYDLNLRPTGGGTGYGRNSVDMCFY 240
Db 181 RRTSETGTGTINTITIKLTQKNGIQIOLPOFKSDARYDLNLRPTGGGTGYGRNSVDMCFY 240
QY 241 DGXSTNSSLEIRFODNPKSDGKFLRKINDDTKEIAYTLISLLAGSLTPNGTSLNI 300
Db 241 DGXSTNSSLEIRFODNPKSDGKFLRKINDDTKEIAYTLISLLAGSLTPNGTSLNI 300
QY 301 ADAASLETNNRITAYTMEPISVPLCWPGRLQDAKVENPEAGQYMGNIIVTFPSSQT 360
Db 301 ADAASLETNNRITAYTMEPISVPLCWPGRLQDAKVENPEAGQYMGNIIVTFPSSQT 360
QY 361 L 361
Db 360 L 360

RESULT 2
US-08-483-101-16
Sequence 16, Application US/08483101
Patent No. 5932715
GENERAL INFORMATION:
APPLICANT: Scott, June R.
APPLICANT: Froehlich, Barbara
TITLE OF INVENTION: CS2 Proteins and Coding Sequences
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee and Winner, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: US
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,101
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Ferber, Donna M.
REGISTRATION NUMBER: 33878
REFERENCE/DOCKET NUMBER: 6-95
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 499-8080
TELEFAX: (303) 499-8089
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 363 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-483-101-16

Query Match 52.4%; Score 988.5; DB 2; Length 363;
Best Local Similarity 55.0%; Pred. No. 1.7e-93;

Matches 204; Conservative 53; Mismatches 93; Indels 21; Gaps 11;
QY 3 KILFIFLFFSSVLETFFAVADKIPGDE--SITNIF-GPR-DRNESSPKHINLNTAY 58
Db 2 KILFIFL-----SIFSAVVSAGRYPEYVGNLTKSQAPRLDSSVSPITNITNVAG 57
QY 59 SESHITLYDMRTFLCLSSHTNLNGACPTSENPSSSVS-GETNITLQFTERSLIKRELOI 117
Db 58 SESHITLYDMRTFLCLSSHTNLNGACPTSENPSSSVS-GETNITLQFTERSLIKRELOI 114
QY 118 KGYKOLLFKNVNPSSG--LTNLNAHFNCKNA-ASGASLYLYTPAGELKNLPFGGIMDAT 174
Db 115 AGMKKPLTWENQSCDFSNLWLNKSSWCGAHGANCTGLNLTPAGELKNLPFGGIMDAT 174
QY 175 LKLRVRRYSET----YGTITNITIKLTQKNGIQIOLPOFKSDARYDLNLRPTGGGTGY 230
Db 175 LKLRVRRYSET----YGTITNITIKLTQKNGIQIOLPOFKSDARYDLNLRPTGGGTGY 230
QY 231 GRNSVDMCFYDGXSTNSSLEIRFODNPKSDGKFLRKINDDTKEIAYTLISLLAGSL 290
Db 234 GRNSVDMCFYDGXSTNSSLEIRFODNPKSDGKFLRKINDDTKEIAYTLISLLAGSL 292
QY 291 TPTNGISLNLADAASLETNNRITAYTMEPISVPLCWPGRLQDAKVENPEAGQYMGNI 350
Db 293 TPTNGISLNLADAASLETNNRITAYTMEPISVPLCWPGRLQDAKVENPEAGQYMGNI 352
QY 351 NVTFTPSSQTL 361
Db 353 KITFTPSSQTL 363

RESULT 3
US-08-483-101-5
Sequence 5, Application US/08483101
Patent No. 5932715
GENERAL INFORMATION:
APPLICANT: Scott, June R.
APPLICANT: Froehlich, Barbara
TITLE OF INVENTION: CS2 Proteins and Coding Sequences
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee and Winner, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: US
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,101
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Ferber, Donna M.
REGISTRATION NUMBER: 33878
REFERENCE/DOCKET NUMBER: 6-95
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 499-8080
TELEFAX: (303) 499-8089
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 364 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-483-101-5

```

Query Match          49.2%; Score 927.5; DB 2; Length 364;
Best Local Similarity 48.1%; Pred. No. 3,3e-87;
Matches 176; Conservative 65; Mismatches 118; Indels 7; Gaps 3;

OY 1 MNKLEFLEFFSSVLETFPAVSADKIPGDESTINIEGPRDRN--ESSPKHNILNHTAY 58
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 LKAVIFVLSMFLCSQVYGQSMHTNVEAGSINKTESIGPIDRSAASYPVHYIFHEHAGY 60

OY 59 SESTILYDRKTFCLSSHNLTNGACPTSENPSSSVSGENITLQTFEKSLLKRELQIK 118
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 NKHSLSFDRMTFLCMSSTDAKSGCAPGENSKSS--OGETNIKILFEKKSLLAKRTLNK 118

OY 119 GYKOLFEKSYNC--PGGLNLSAHFNCNKAASGASLYIYPAGELKNLPGGIMPATL 175
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 119 GYKRFLESPRCIHVDKMLNSHTVACVGSFTRGVDFLYIIPGELDGLITGIMPATL 178

OY 176 KLRVRRYSETGYTYTINIKTLKDNQIOWLPQFKSARVDLNLAPTGSGTYIGNSV 235
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 179 ELKRVKRYDYNHGKYKYNITVDLTKDGNIQVMPKPFHSODRIDNLNRPENKSGSNVL 238

OY 236 DMCFTYDGYSTNSSLEIRFODNNPKSDGKFLYLRKINDTKEIATVLSLLAGSLTPNG 295
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 239 EMLCYDGYSTHSSQSIEMKRFDDSDGTGNEYNLKTGEPKLPKYLKSLILGREFYPNG 298

OY 296 TSLNIADAASLETNMRITAVTMEISVPLCWPGRQLQDAKVENPBAQYMGNIYTF 355
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 299 EAFINDTSLFTNMNRIKSVSLPOLSIPLVLCWPMANLTFMSELNPEAGEISGLINVT 358

OY 356 PSSQTL 361
    : : : : :
Db 359 PSSSSL 364

RESULT 4
US-08-617-697-10
; Sequence 10, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Matzare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617, 697
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 05-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-557
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 10:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1600 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-617-697-10

Query Match          6.2%; Score 116.5; DB 2; Length 1600;
Best Local Similarity 21.6%; Pred. No. 0.014;
Matches 74; Conservative 60; Mismatches 130; Indels 79; Gaps 19;

OY 48 HNLLNHTAYSSHTLYDMLTFCLSSHNLTNGACPTSENPSSS--SVSGETITL 102
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 816 YNEYSKH--AINSSHNL-----TILGNAVTLG-----ENSSSITGINITNKAVAL 862

OY 103 QF-----TEKSLIKRELQIKGYKOLFEKSYNCPGTLT-----NSAHFNCN-- 144
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 863 QADTSNNTGLKRRKTLTNGINISVEGNLSLGAMNINIGNSIAMEDTFKGEASDNLITG 922

OY 145 ---KNAASGASLYIYPAGELKNLPGGIMPATLKLVRKRYSETYGT--YTINITKLT 200
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 923 TFTNNGTANINIKGVYALGDIINNK--GG-----LNTTNASGTOKTIINGNT--NE 970

OY 201 KGNIQIQLPQFKSARVDL--NLAPTGSGTYIGNSVDMCFYDGYSTNSSSLEIRQ--- 255
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 971 KGDINT--KNIKDAEIQIGNISQKEGNTLTISDKVNT-----TNQITKAGVGR 1021

OY 256 -DNNPKSDGKFLYLRKINDTKEIATVLSLLAG---KSLPTNGTSLNIADAASLETNWN 311
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1022 SDSSEANANLTIQ-----TKELKLAGDLNLSGPNKAEITAKNSDGLTINASGNNDAK 1076

OY 312 RITAVTMEISVPLCWPGR--RLQDAKVENPBAQYMGNIYTF 353
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1077 K--VFEDKVKDSKISTGDHNVTLNLSKVTSNGSSNAGNDNST 1116

RESULT 5
US-08-362-525-2
; Sequence 2, Application US/08362525
; Patent No. 6027910
; GENERAL INFORMATION:
; APPLICANT: KILS, FRANCISCUS M.
; APPLICANT: SCHREIDER, MARTEN P.
; APPLICANT: TOSCHKA, HOLSER Y.
; APPLICANT: VERRIPS, CORNELIS T.
; TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ENZYMES TO THE
; TITLE OF INVENTION: CELL WALL OF A MICROBIAL CELL BY PRODUCING A FUSION
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBAY & CUSHMAN, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362,525
; FILING DATE: 04-JAN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 92202080.5
; FILING DATE: 08-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 92203899.7
; FILING DATE: 14-DEC-1992
; PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: PCT/EP93/01763
FILING DATE: 07-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N.
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 213389/T7020(V)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 861-3000
TELEFAX: (202) 822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 650 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-362-525-2

Query Match 6.0%; Score 114; DB 3; Length 650;
Best local Similarity 20.9%; Pred. No. 0.0057;
Matches 98; Conservative 52; Mismatches 147; Indels 172; Gaps 23;

QY 4 ILPIFLFESSVL-----FTFA-----VSADKIPG-----DESTINIGPPDRNE--- 43
DB 8 ILWLFSLALASAININDIFPSNLEITPLANKOPDQGWATPFDSIADASSIREGDEFTL 67
QY 44 SSP---KHNILNHNHTA-----YSESHLYDRMTPLC-----LSNHTLN 80
DB 68 SMHVYRIKILNSOTATITSLADGTEFAKCYVSOQAAYLTENTTFTCTAONDLSSVTFID 127
QY 81 GACPTSENPSSSVSGETNITTOFTTEKRSLIKRELOIKGYKOLLFKSVNC-PSGLTLNSA 139
DB 128 GSITFSINFDGSSSYEYL-----ENAKFKSGPMLVKGNQMSDVVNNDPAAFTEVNF 182
QY 140 HFQCNKNAASGASLYL--YIPAGELKNLPPGGIWDATLKLKVRKRRISSETYGTITNITIK 197
DB 183 HSGRSRGYGSFESYHLGMCPCPNFYF---LGG---TEKI-----DIDSSNNVNDL- 225
QY 198 LTKGNIGI-----MLPQKSDARVDLNLPTGGGTYIGRN---SYDMCFYDGS--- 242
DB 226 --DCSSVQVSSNDPNDMPFOSYNDTNDVY-----TCFSGNLTITLDEKLYDEML 275
QY 243 -----YSTNSSSL-IRPDNN--PKSDGKF 265
DB 276 WYNALGSLPANVNTIDHALEFYTCLODTIANTTYATQFSTRREIYVQGNLGTASAKSS 335
QY 266 YLRKINDTKET---AYTL-----LILAKSLPTPTNGSLNADAA 304
DB 336 FISTTTTDLTISMTSAYSTGSIETVENGTSTSEVISHVVTSTKLSPTATSLTIAOTS 395
QY 305 SLETFNNM-----RTATVMPETISVPVLCMPGR 332
DB 396 IYSTDNITVGTIDHTTSEVISDETISRETASTVAAPTSTGWTGAM 444

RESULT 6
US-08-728-470-10
Sequence 10, Application us/08728470
Patent No. 5928651
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mathare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,470
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-633
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0810
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1529 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-728-470-10

Query Match 5.9%; Score 111.5; DB 2; Length 1529;
Best local Similarity 20.7%; Pred. No. 0.041;
Matches 72; Conservative 62; Mismatches 127; Indels 87; Gaps 18;

QY 48 HNILNHNITVASESHLYDRMTPLCLSHNLTNGACPTSENPSS-----SVSGETNITL 102
DB 744 YNEYSKH--AINSSHL-----TILGNAVTLG-----ENSSSTITGNINITNKANVTL 790
QY 103 QF-----TEKRSLIKRELOIKGYKOLLFKSVNCPGSLTLNSAHFNCKNAASGASLY 154
DB 791 QADTSNMGKLRKRLTLGIVSEGNLSLGANANIVGNLSI--ARDSTFKGAS----- 843
QY 155 LYIPAGELKNLPPGGIWDATLKLKVRKRRISSETYGTITNITIKL-----INTIKL----- 199
DB 844 -----DNLNTTGTFTNGTANINIKOGVYVLOGDINNKGGLNITTNASTGTOKTIING 895
QY 200 ---DKNGIOIWLPPQKSDARVDL--NLPTGGGTYIGRNSVDMCFYDGSYSTNSSLEIR 253
DB 896 NITNEKGDINT--KNIKADAEIOIGNISOKESNLITSSDKYNI-----INQITIKAG 946
QY 254 FQ---DNNKSDGKFLYLRKINDTKETIAYTSLLAG---KSLPTNGTSLNADAA 306
DB 947 VEGGRSDSSEANANLTIQ---TKELKLAGDLINISFNKAEITFAKNSDGLTIGNASGG 1001
QY 307 ETWMNKITAVTMEISVPVLCMPG-RLQIDAKYENPEAQYMKMNIVT 353
DB 1002 NADAKR---VTFDKVADSKISTDGHNVTLINSEVKTSGSSMAGNDNST 1046

RESULT 7
US-08-719-641-10
Sequence 10, Application us/08719641
Patent No. 6218141
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/719,641
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Belkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-625
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1529 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-719-641-10

Query Match 5.9%; Score 111.5; DB 4; Length 1529;
Best Local Similarity 20.7%; Pred. No. 0.041;
Matches 72; Conservative 62; Mismatches 127; Indels 87; Gaps 16;

QY 48 HNLNHTAVSESHLYDRMTFLCLSSHNTLNGACPTSENPSSS-----SVSGFTNITL 102
DB 744 YNEYSKH--AINSSHNL-----TILGCVNVLG-----ENSSSITGINTITNKANVTL 790
QY 103 QF-----TEKRLIRRELQIKYKQLLFKRSVNCPSGLTINSAHFNCNKAAGASLY 154
DB 791 QADTNSNMTGKLKRLTLGNISVEGNLSLTGANANIVGNLSI--AEDSTFKGAS----- 843
QY 155 LYIPAGELKLNLPFGGIW--DATLKLVRKRYSEYGTYT---INITIKLT----- 199
DB 844 -----DNLNTITFTFNNGTANININIGGVKKLOGDINNKGGLTITTNASTGTITING 895
QY 200 ----DKGNIQIWLPOFKSDARVDL--NLPTGGGTYIGRNSVDMCFYDGYSTNSSLEIR 253
DB 896 NITNEKGLNLI--KNIKADAELQIGCNISQKNGNLTISSDKVNI-----TNOQTTIKAG 946
QY 254 FO-----DNMPSDGFYLRKINDTKELAYTILSLLAG---KSLPTNGTSLINTADAASL 306
DB 947 VEGGRSDSESEANLTIQ-----TKELKLAGDLNISGFENKAETIAKKNGSDDLTIIGNASG 1001
QY 307 ETNNMRTIAYTMPEISVPLCMPG--RLQDLAKVENPEAGQYVGNINVT 353
DB 1002 NADAK-----VTFDKYKDSKISTDGHNTVNLSEVKTSGNSSNAGNDNST 1046

RESULT 8
US-08-737-716-13
; Sequence 13, Application US/08737716

Patent No. 5955258
GENERAL INFORMATION:
APPLICANT: GILBE BUIST
APPLICANT: GERARD VENEMA
APPLICANT: JAN KOK
TITLE OF INVENTION: Process for the lysis of a culture of lactic
TITLE OF INVENTION: acid bacteria by means of a lysis, and uses of the resultin
TITLE OF INVENTION: lysed culture.
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pillsbury Madison & Sutro, L.L.P.
STREET: 1100 New York Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/737,716
FILING DATE: 22-APR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/NL95/00170
FILING DATE: 12-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 94201353.3
FILING DATE: 12-MAY-1994
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 671 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Streptococcus faecalis
IMMEDIATE SOURCE:
CLONE: Fig.5a (S. faecalis)
US-08-737-716-13

Query Match 5.6%; Score 106; DB 2; Length 671;
Best Local Similarity 23.6%; Pred. No. 0.04;
Matches 73; Conservative 39; Mismatches 103; Indels 94; Gaps 17;

QY 41 RNESSPKHNT-LNNHTIYAVSESHLYDRMTFLCLSSHNTLNGACPTSENPSSSV----- 94
DB 309 RYATDPSPYNAKLNINVTAY--NLTYDTPSSGNTGGTVNGTGSNNQSGTNNYYTVK 366
QY 95 SGET--NTTLOF-----TEKRL--IKRELQIKYKQLLFKRSVNCPSGLTINSAHFNCNKN 146
DB 367 SODTLNKIAAQAQGYVAVANLRKSWNGISGDIIFYGOKLIYAKKA--SGNTGSGNGSGNNN 423
QY 147 AAGASLYLYIPAGELKLNLPFGGIW--DATLKLVRKRYSEYGTYTINITIKLTDKGNIQI 206
DB 424 -QSGTNYTYTVKSGDTLN-----KIAAQY-----VTY-----ANLRS 455
QY 207 WLPOFKSDARVDLNL-----RPTGGGTYIGRNSVDMCFYDGYSTNSSLEIRFODNN 258
DB 456 W-----NNGISGLILVVGOKLIYKKTSGNT-----GSSNGS-----NNN 491
QY 259 PKSDGKFLYLRKINDTKELA--YTLS-----LLIAGSLPTNGTSLINTADA 303
DB 492 QSGTNYTYTVKSGDTLNKIAAQAQGYVAVANLRKSWNGISGDLIFAGOKITTYKKTSGNTGGS 551
QY 304 ASLETNNMR 312
DB 552 SNGSGNNNO 560

```

RESULT 9
US-08-728-470-9
: Sequence 9, Application US/08728470
: Patent No. 5928651
: GENERAL INFORMATION:
: APPLICANT: Barenkamp, Stephen J
: TITLE OF INVENTION: High Molecular Weight Surface Proteins
: TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Mattare, Ltd.
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: STREET: Bldg. 1
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/728,470
: FILING DATE:
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/302,832
: FILING DATE: 16-MAR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US PCT/US93/02166
: FILING DATE: 16-MAR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 9205704.1
: FILING DATE: 16-MAR-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Belkstresser, Jerry W
: REGISTRATION NUMBER: 22,651
: REFERENCE/DOCKET NUMBER: 1038-633
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0813
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1338 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-728-470-9

```

Query Match	5.4%;	Score 101;	DB 2;	Length 138;
Best Local Similarity	21.5%;	Pred. No. 0.4;		
Matches	90;	Conservative	52;	Mismatches 172; Indels 104; Gaps 18;
OY	21	VSADKIPDESTITNFGPRDNESPKNH-LNNHI-TAYSEHTLYDRMFCLCS----	74	
		: :	: : : : : : : :	
Db	817	VFEDKV-----KDSKISIDGHNVTLSNVEVKTSGSSNAGNNDNSTGLTISAKDV	864	
OY	75	-SHNTLN-----GACPTSENPSSSVSGETNTLLO-----FTEKR	108	
		: : : : : : : :		
Db	865	TVNNNVTSHKRTINISAAGNVTTKRGCTIINATYGSVEVYTAONGTICKNTTSQNVYTAIE	924	
OY	109	SLIKRELOI---KQYKOLLEKSVCPGGLTINSAHFENCKNKAASGAST-----	153	
		: : : : : : : : : : : : : : : : : :		
Db	925	NLVYTEMVNATSGTVNISTKGTGKIGKIGESTSONVNI---TASGNTLKVSNITGQDVT	981	
OY	154	-----YAIYPAGELEKNLPFGGIMDATLKLVRKRRYSERFYGTNTITKLDKGIQIOM	207	
		: : : : : : : : : : : : : : : : :		
Db	982	VTADAGALTITAGSTISATNGTANNTTTGDLNGKVESSSGSVTLVATGATILAVGNISGN	1041	
OY	208	LPQFSDAKRVLDLNPFGGGITYIGNNSVDMCFYDGTISTNSSLLEIRFQDNN---PKSDCK	264	

```

Db      1042  TTTTADSG---KLTSTVSGTNGTNSVTT-----SSQSGDEGTISGNTVYNTASTGD 1092

Oy      265  FYL-RKINDOTKEATYLS-----LLAGKSITPPIGNSLIMADAALEINMN--RTA 315
      :      :      :      :      :      :      :      :      :      :
Db      1093  LTTIINSAKVFAKNGAATLTFAESKLTPTQTGSSITSSNGOTTLTADSSIAGINAANYTL 1152

Oy      316  VTMEPEISV---PYLCPGPILOT---DAVENPEAG-----QYMNINWTFPPSS 358
      :      :      :      :      :      :      :      :      :      :
Db      1153  NTTGTGTLTTTGDSKINATSGTTLTNADAKLDGASASDRVYVATNTAASSGNVTAKTSS 1210

```

```

10 RESULT
11 US-08-719-641-9
12 ; Sequence 9, Application US/08719641
13 ; Patent No. 6218141
14 ;
15 ; GENERAL INFORMATION:
16 ; APPLICANT: Barenkamp, Stephen J
17 ; TITLE OF INVENTION: High Molecular Weight Surface Proteins
18 ; TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
19 ; NUMBER OF SEQUENCES: 10
20 ;
21 ; CORRESPONDENCE ADDRESS:
22 ; ADDRESSEE: Shoemaker and Mattare, Ltd.
23 ; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
24 ; STREET: Bldg. 1
25 ; CITY: Arlington
26 ; STATE: Virginia
27 ; COUNTRY: U.S.A.
28 ; ZIP: 22202-0286
29 ;
30 ; COMPUTER READABLE FORM:
31 ; MEDIUM TYPE: Floppy disk
32 ;
33 ; COMPUTER: IBM PC compatible
34 ; OPERATING SYSTEM: PC-DOS/MS-DOS
35 ; SOFTWARE: Patentln Release #1.0, Version #1.30
36 ;
37 ; CURRENT APPLICATION DATA:
38 ; APPLICATION NUMBER: US/08/719,641
39 ; FILING DATE:
40 ;
41 ; CLASSIFICATION: 530
42 ;
43 ; PRIOR APPLICATION DATA:
44 ; APPLICATION NUMBER: US 08/302,832
45 ; FILING DATE: 16-SEP-1994
46 ;
47 ; PRIOR APPLICATION DATA:
48 ; APPLICATION NUMBER: US 08/302,832
49 ; FILING DATE: 16-MAR-1993
50 ;
51 ; PRIOR APPLICATION DATA:
52 ; APPLICATION NUMBER: GB 9205704.1
53 ; FILING DATE: 16-MAR-1992
54 ;
55 ; ATTORNEY/AGENT INFORMATION:
56 ; NAME: Belkstresser, Jerry W
57 ; REGISTRATION NUMBER: 22,651
58 ; REFERENCE/DOCKET NUMBER: 1038-625
59 ;
60 ; TELECOMMUNICATION INFORMATION:
61 ; TELEPHONE: (703) 415-0810
62 ; TELEFAX: (703) 415-0813
63 ;
64 ; INFORMATION FOR SEQ ID NO: 9:
65 ;
66 ; SEQUENCE CHARACTERISTICS:
67 ; LENGTH: 1338 amino acids
68 ; TYPE: amino acid
69 ; STRANDEDNESS: single
70 ; TOPOLOGY: linear
71 ;
72 ;
73 ;
74 ;
75 ;
76 ;
77 ;
78 ;
79 ;
80 ;
81 ;
82 ;
83 ;
84 ;
85 ;
86 ;
87 ;
88 ;
89 ;
90 ;
91 ;
92 ;
93 ;
94 ;
95 ;
96 ;
97 ;
98 ;
99 ;
100 ;
101 ;
102 ;
103 ;
104 ;
105 ;
106 ;
107 ;
108 ;
109 ;
110 ;
111 ;
112 ;
113 ;
114 ;
115 ;
116 ;
117 ;
118 ;
119 ;
120 ;
121 ;
122 ;
123 ;
124 ;
125 ;
126 ;
127 ;
128 ;
129 ;
130 ;
131 ;
132 ;
133 ;
134 ;
135 ;
136 ;
137 ;
138 ;
139 ;
140 ;
141 ;
142 ;
143 ;
144 ;
145 ;
146 ;
147 ;
148 ;
149 ;
150 ;
151 ;
152 ;
153 ;
154 ;
155 ;
156 ;
157 ;
158 ;
159 ;
160 ;
161 ;
162 ;
163 ;
164 ;
165 ;
166 ;
167 ;
168 ;
169 ;
170 ;
171 ;
172 ;
173 ;
174 ;
175 ;
176 ;
177 ;
178 ;
179 ;
180 ;
181 ;
182 ;
183 ;
184 ;
185 ;
186 ;
187 ;
188 ;
189 ;
190 ;
191 ;
192 ;
193 ;
194 ;
195 ;
196 ;
197 ;
198 ;
199 ;
200 ;
201 ;
202 ;
203 ;
204 ;
205 ;
206 ;
207 ;
208 ;
209 ;
210 ;
211 ;
212 ;
213 ;
214 ;
215 ;
216 ;
217 ;
218 ;
219 ;
220 ;
221 ;
222 ;
223 ;
224 ;
225 ;
226 ;
227 ;
228 ;
229 ;
230 ;
231 ;
232 ;
233 ;
234 ;
235 ;
236 ;
237 ;
238 ;
239 ;
240 ;
241 ;
242 ;
243 ;
244 ;
245 ;
246 ;
247 ;
248 ;
249 ;
250 ;
251 ;
252 ;
253 ;
254 ;
255 ;
256 ;
257 ;
258 ;
259 ;
260 ;
261 ;
262 ;
263 ;
264 ;
265 ;
266 ;
267 ;
268 ;
269 ;
270 ;
271 ;
272 ;
273 ;
274 ;
275 ;
276 ;
277 ;
278 ;
279 ;
280 ;
281 ;
282 ;
283 ;
284 ;
285 ;
286 ;
287 ;
288 ;
289 ;
290 ;
291 ;
292 ;
293 ;
294 ;
295 ;
296 ;
297 ;
298 ;
299 ;
300 ;
301 ;
302 ;
303 ;
304 ;
305 ;
306 ;
307 ;
308 ;
309 ;
310 ;
311 ;
312 ;
313 ;
314 ;
315 ;
316 ;
317 ;
318 ;
319 ;
320 ;
321 ;
322 ;
323 ;
324 ;
325 ;
326 ;
327 ;
328 ;
329 ;
330 ;
331 ;
332 ;
333 ;
334 ;
335 ;
336 ;
337 ;
338 ;
339 ;
340 ;
341 ;
342 ;
343 ;
344 ;
345 ;
346 ;
347 ;
348 ;
349 ;
350 ;
351 ;
352 ;
353 ;
354 ;
355 ;
356 ;
357 ;
358 ;
359 ;
360 ;
361 ;
362 ;
363 ;
364 ;
365 ;
366 ;
367 ;
368 ;
369 ;
370 ;
371 ;
372 ;
373 ;
374 ;
375 ;
376 ;
377 ;
378 ;
379 ;
380 ;
381 ;
382 ;
383 ;
384 ;
385 ;
386 ;
387 ;
388 ;
389 ;
390 ;
391 ;
392 ;
393 ;
394 ;
395 ;
396 ;
397 ;
398 ;
399 ;
400 ;
401 ;
402 ;
403 ;
404 ;
405 ;
406 ;
407 ;
408 ;
409 ;
410 ;
411 ;
412 ;
413 ;
414 ;
415 ;
416 ;
417 ;
418 ;
419 ;
420 ;
421 ;
422 ;
423 ;
424 ;
425 ;
426 ;
427 ;
428 ;
429 ;
430 ;
431 ;
432 ;
433 ;
434 ;
435 ;
436 ;
437 ;
438 ;
439 ;
440 ;
441 ;
442 ;
443 ;
444 ;
445 ;
446 ;
447 ;
448 ;
449 ;
450 ;
451 ;
452 ;
453 ;
454 ;
455 ;
456 ;
457 ;
458 ;
459 ;
460 ;
461 ;
462 ;
463 ;
464 ;
465 ;
466 ;
467 ;
468 ;
469 ;
470 ;
471 ;
472 ;
473 ;
474 ;
475 ;
476 ;
477 ;
478 ;
479 ;
480 ;
481 ;
482 ;
483 ;
484 ;
485 ;
486 ;
487 ;
488 ;
489 ;
490 ;
491 ;
492 ;
493 ;
494 ;
495 ;
496 ;
497 ;
498 ;
499 ;
500 ;
501 ;
502 ;
503 ;
504 ;
505 ;
506 ;
507 ;
508 ;
509 ;
510 ;
511 ;
512 ;
513 ;
514 ;
515 ;
516 ;
517 ;
518 ;
519 ;
520 ;
521 ;
522 ;
523 ;
524 ;
525 ;
526 ;
527 ;
528 ;
529 ;
530 ;
531 ;
532 ;
533 ;
534 ;
535 ;
536 ;
537 ;
538 ;
539 ;
540 ;
541 ;
542 ;
543 ;
544 ;
545 ;
546 ;
547 ;
548 ;
549 ;
550 ;
551 ;
552 ;
553 ;
554 ;
555 ;
556 ;
557 ;
558 ;
559 ;
560 ;
561 ;
562 ;
563 ;
564 ;
565 ;
566 ;
567 ;
568 ;
569 ;
570 ;
571 ;
572 ;
573 ;
574 ;
575 ;
576 ;
577 ;
578 ;
579 ;
580 ;
581 ;
582 ;
583 ;
584 ;
585 ;
586 ;
587 ;
588 ;
589 ;
590 ;
591 ;
592 ;
593 ;
594 ;
595 ;
596 ;
597 ;
598 ;
599 ;
600 ;
601 ;
602 ;
603 ;
604 ;
605 ;
606 ;
607 ;
608 ;
609 ;
610 ;
611 ;
612 ;
613 ;
614 ;
615 ;
616 ;
617 ;
618 ;
619 ;
620 ;
621 ;
622 ;
623 ;
624 ;
625 ;
626 ;
627 ;
628 ;
629 ;
630 ;
631 ;
632 ;
633 ;
634 ;
635 ;
636 ;
637 ;
638 ;
639 ;
640 ;
641 ;
642 ;
643 ;
644 ;
645 ;
646 ;
647 ;
648 ;
649 ;
650 ;
651 ;
652 ;
653 ;
654 ;
655 ;
656 ;
657 ;
658 ;
659 ;
660 ;
661 ;
662 ;
663 ;
664 ;
665 ;
666 ;
667 ;
668 ;
669 ;
670 ;
671 ;
672 ;
673 ;
674 ;
675 ;
676 ;
677 ;
678 ;
679 ;
680 ;
681 ;
682 ;
683 ;
684 ;
685 ;
686 ;
687 ;
688 ;
689 ;
690 ;
691 ;
692 ;
693 ;
694 ;
695 ;
696 ;
697 ;
698 ;
699 ;
700 ;
701 ;
702 ;
703 ;
704 ;
705 ;
706 ;
707 ;
708 ;
709 ;
710 ;
711 ;
712 ;
713 ;
714 ;
715 ;
716 ;
717 ;
718 ;
719 ;
720 ;
721 ;
722 ;
723 ;
724 ;
725 ;
726 ;
727 ;
728 ;
729 ;
730 ;
731 ;
732 ;
733 ;
734 ;
735 ;
736 ;
737 ;
738 ;
739 ;
740 ;
741 ;
742 ;
743 ;
744 ;
745 ;
746 ;
747 ;
748 ;
749 ;
750 ;
751 ;
752 ;
753 ;
7
```

[illegible]

Db 865 TVNNVNTSHKNTINISAAGNVTTKEGTTINATGSEVTAQNGTIKGNITSQNVYATATE 924
QY 109 SLIKREIQI-----KGYKQLFKSVNCPGSLTNSAHFNCKNMAASGSL----- 153
Db 925 NLVTTENAIVNATSGVYNIISTKGTGDIKGIESTSGNVNI---TASGNTLVNSITGDDVT 981
QY 154 -----YLYIPAGELKMLPFGGIMDATLKLVRKRRYSEYGYTYTINTIKLTDKGNQIW 207
Db 982 VTADAGALTTTASGTSISATGNNITTKTDINGKEVSSGSLVATGATLAVGNISGN 1041
QY 208 LPOFKSDARVDMLRPTGGGTYIGRNSVDMCFYDGYSTNSSLEIRPDNN---PKSDGK 264
Db 1042 TVTITADSG---KLTSTVGSTINGTNSVTT-----SSQSGDIEGTISGNTVNVYATSGD 1092
QY 265 FYL-RKINDTKELIATLS-----LLAGKSLPTNGTSLNIAAASLETNNN--RITA 315
Db 1093 LITIGNSAKVEAKNGAATLPRESKLTPTOTGSSITSSNGQTTTLAKSSISGINIAAIVTL 1152
QY 316 VTMPETISV---PYLCMPGRLOL---DAKYENPEAG-----QYGNINVTFTPS 358
Db 1153 NTGTGTTTGDSKINATSGTLTINAKDAKLDGAASGDRIVVATNATNAGSGNVYAKTSS 1210

RESULT 11
US-08-617-697-9
; Sequence 9, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Matzare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,697
; FILING DATE: 01-Apr-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 05-Oct-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-557
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0810
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1599 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-617-697-9

Query Match 5.4%; Score 101; DB 2; Length 1599;
Best Local Similarity 21.5%; Pred. No. 0.54;

Matches 90; Conservative 52; Mismatches 172; Indels 104; Gaps 18;
QY 21 VSADKIPGDESTINIFEPORRNESSPKHN1-LNNHT-TAISSEHTLYDRATFLCLS---- 74
Db 1077 VTEDKV-----KSKISTDGHNVTLNSEVTSNGSSNAGNDNSTGLTISAKDV 1124
QY 75 -----SHNTLN-----GACPTSENPSSSSVSGETNITLO-----FEKR 108
Db 1125 TVNNVNTSHKNTINISAAGNVTTKEGTTINATGSEVTAQNGTIKGNITSQNVYATATE 1184
QY 109 SLIKREIQI-----KGYKQLFKSVNCPGSLTNSAHFNCKNMAASGSL----- 153
Db 1185 NLVTTENAIVNATSGVYNIISTKGTGDIKGIESTSGNVNI---TASGNTLVNSITGDDVT 1241
QY 154 -----YLYIPAGELKMLPFGGIMDATLKLVRKRRYSEYGYTYTINTIKLTDKGNQIW 207
Db 1242 VTADAGALTTTASGTSISATGNNITTKTDINGKEVSSGSLVATGATLAVGNISGN 1301
QY 208 LPOFKSDARVDMLRPTGGGTYIGRNSVDMCFYDGYSTNSSLEIRPDNN---PKSDGK 264
Db 1302 TVTITADSG---KLTSTVGSTINGTNSVTT-----SSQSGDIEGTISGNTVNVYATSGD 1352
QY 265 FYL-RKINDTKELIATLS-----LLAGKSLPTNGTSLNIAAASLETNNN--RITA 315
Db 1353 LITIGNSAKVEAKNGAATLPRESKLTPTOTGSSITSSNGQTTTLAKSSISGINIAAIVTL 1412
QY 316 VTMPETISV---PYLCMPGRLOL---DAKYENPEAG-----QYGNINVTFTPS 358
Db 1413 NTGTGTTTGDSKINATSGTLTINAKDAKLDGAASGDRIVVATNATNAGSGNVYAKTSS 1470

RESULT 12
US-08-296-791-6
; Sequence 6, Application US/08296791
; Patent No. 6245337
; GENERAL INFORMATION:
; APPLICANT: St. Gene III, Joseph W.
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/296,791
; FILING DATE: 25-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1848 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; US-08-296-791-6

Query Match 5.2%; Score 99; DB 4; Length 1848;
Best Local Similarity 23.4%; Pred. No. 1.1;
Matches 76; Conservative 56; Mismatches 127; Indels 66; Gaps 20;

QY 19 FAVSADKIGDSEITNIPFPRDRNESSPKHNLIINHTAYSES--HTLDRMTFLCLSS- 75
DB 753 FRAITMNVGNASLVS-----GRNVA---NITSN--ITASNNAGVHIGKTDGDTVCVRSD 802
QY 76 -----HNT-LNGACPTSENPSSSVSGEFTNT--LQFTEKRSILKRELQIKGYKQLF 125
DB 803 YGVYVCHNSNLSEKALNFNP--TILRGVNLTEASTLTKANLFGTIQSIGTQVNL 860
QY 126 KSYNCPGSLTNSAHFNCK--NAAGSASLYLYIPAGELKNLPFGIWDATLKLVRKRRYS 184
DB 861 KE-NSHMLTGN--NVQMLNTNG---HILMAQNDAN-----KV 895
QY 185 EYVGYTITITIKLDKNGIQIWLPOFKSDA--RYDLNLPRPTGGCTY-----IGR-NSVDM 237
DB 896 TYTNLTIVN---SLSGNSGFYTWDFTNKSNKVAVNKSATGFTLQVADKTGEPHNEL 952
QY 238 CFYDGYSTNSSLEIRFQDNN--PKSDGFEYLRKINDTKEIAYTSLLAGKSLPTNGT 296
DB 953 TLFDAASNATRNNEVLVLANGSVDRGAMKYLKLRNVNG--RYDLYNPEVEKRNQTVDTTNT 1010
QY 297 SLN--TADAASLETNNRITAVTMP 319
DB 1011 TPNDIQADAPASQSNNEETIARVETP 1035

RESULT 13

PCT-US95-10661A-6
; Sequence 6, Application PC/TUS9510661A
; GENERAL INFORMATION:
; APPLICANT: Washington University, et al.
; TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10661A
; FILING DATE: 16-AUG-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,791
; FILING DATE: 25-AUG-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Treacartn, Richard F.
; REGISTRATION NUMBER: 31,801
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELECOMMUNICATION INFORMATION:
; REFERENCE/DOCKET NUMBER: FP-59941/RFT
; TELEPHONE: (415) 781-1989
; TELEFAX: 910 277299
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1848 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; PCT-US95-10661A-6

Query Match 5.2%; Score 99; DB 5; Length 1848;

Best Local Similarity 23.4%; Pred. No. 1.1;
Matches 76; Conservative 56; Mismatches 127; Indels 66; Gaps 20;

QY 19 FAVSADKIGDSEITNIPFPRDRNESSPKHNLIINHTAYSES--HTLDRMTFLCLSS- 75
DB 753 FRAITMNVGNASLVS-----GRNVA---NITSN--ITASNNAGVHIGKTDGDTVCVRSD 802
QY 76 -----HNT-LNGACPTSENPSSSVSGEFTNT--LQFTEKRSILKRELQIKGYKQLF 125
DB 803 YGVYVCHNSNLSEKALNFNP--TILRGVNLTEASTLTKANLFGTIQSIGTQVNL 860
QY 126 KSYNCPGSLTNSAHFNCK--NAAGSASLYLYIPAGELKNLPFGIWDATLKLVRKRRYS 184
DB 861 KE-NSHMLTGN--NVQMLNTNG---HILMAQNDAN-----KV 895
QY 185 EYVGYTITITIKLDKNGIQIWLPOFKSDA--RYDLNLPRPTGGCTY-----IGR-NSVDM 237
DB 896 TYTNLTIVN---SLSGNSGFYTWDFTNKSNKVAVNKSATGFTLQVADKTGEPHNEL 952
QY 238 CFYDGYSTNSSLEIRFQDNN--PKSDGFEYLRKINDTKEIAYTSLLAGKSLPTNGT 296
DB 953 TLFDAASNATRNNEVLVLANGSVDRGAMKYLKLRNVNG--RYDLYNPEVEKRNQTVDTTNT 1010
QY 297 SLN--TADAASLETNNRITAVTMP 319
DB 1011 TPNDIQADAPASQSNNEETIARVETP 1035

RESULT 14

US-08-409-995-4
; Sequence 4, Application US/08409995
; Patent No. 5646259
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen I.
; TITLE OF INVENTION: Haemophilus Adhesion Proteins
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/409,995
; FILING DATE: 24-MAR-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-61053/RFT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEFAX: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1912 amino acids
; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; US-08-409-995-4

Query Match 5.0%; Score 94.5; DB 1; Length 1912;
Best Local Similarity 19.7%; Pred. No. 3.4;
Matches 61; Conservative 47; Mismatches 117; Indels 85; Gaps 14;


```

0Y 78 TLNACPTSENPSSVSGETINTLOFTEKRSILIKELQIKQKOLLEFSVNC-----PSG 133
Db 805 TIGENTFP-----GGTATPVPVNT-STADGLNPAKETADASSGSKNYELGIAITTLTESPA 859
0Y 134 LTLNAPHCNKNNA---ASGASLYLIPA-----GELKMLPEGGIWDATLKLVRKRRSE 185
Db 860 -GAKSHVDLNDVATKKSMASTIEDVLRAGMNIQSGNNDVYATYDY-----NFTDD 912
0Y 186 TYGTYTINITIKLTDK-----NIQIMLPQFSKSDARVDNL-----RPTG 225
Db 913 STGTYTTLVYQKADGKADYKIGAKTSYIKDHNGKLFTEKDLKDANNGATSEDDGKDTG 972
0Y 226 GGTITIGRNSVDMCFDGY-----STNSSLEIRFODNN-----258
Db 973 TGLTYATVTDIVAVNKSGBRVTEGATAETGATVAVNAGNAETVTSCTSVMKNGMTATATV 1032
0Y 259 PKSDGKRYLR-----KINDTEIAYTLSTL-LLAASKLPTNGTSLN-----IADA 304
Db 1033 SKDNGNINVKRIDVANGDLKIGDKRKIVADPTTLTVTGKRSVPAGANSVNNKKLVAAE 1092
0Y 305 SLETFNNRIT 314
Db 1093 GLATLALNLS 1102

```

RESULT 15
US-08-685-467-4
; Sequence 4, Application US/08685467
; Patent No. 6060059
; GENERAL INFORMATION:
RECORDING OF INFORMATION

APPLICANT: St. Geme III, Joseph W.
 APPLICANT: Barenkamp, Stephen J.
 TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
 NUMBER OF SEQUENCES: 6
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Fleh, Hohbach, Test, Albritton & Herbert
 STREET: Four Embarcadero Center, Suite 3400
 CITY: San Francisco
 STATE: California
 COUNTRY: United States
 ZIP: 94111-4187
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/685,467
 FILING DATE: 22-JUL-1996
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/409,995
 FILING DATE: 24-MAR-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Silva, Robin M.
 REGISTRATION NUMBER: 38,304
 REFERENCE/DOCKET NUMBER: A-61053-2/RFT/RMS
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 781-1989
 TELEFAX: (415) 398-3249
 TELEX: 910 277299
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1912 amino acids
 TYPE: amino acid
 STRANDEDNESS: unknown
 TOPOLOGY: unknown
 MOLECULE TYPE: protein
 US-08-685-467-4

[illegible]

Search completed: July 1, 2002, 11:24:30
Job time: 302 sec

```
QY      305 SLETTNWRIT 314
          | | | : :
Db      1093 GLATALLNLS 1102
```

```

Y      259  PMSDGEFLKR-----KINDTKELATVLSL-LIAGSLPTNGTSLN-----IADA 304
Y      1033  SFDNGINNVKKYDVNVGDGLKIGDDKKIADPTTTLTVGKVSVPAGANSVNNKKLVAAE 10923
b      Y      305  SLEETMNNRTI 314
      b      1093  GIATALNNLS 1102

RESULT 15
S-08-685-467-4
Sequence No 4, Application US/08685467
Patent No. 6060050

```

SEQUENCE 4, Application US/08685467
Patent No. 6060059
GENERAL INFORMATION:

APPLICANT: St. Gene III Joseph W.
 APPLICANT: Barekcamp, Stephen J.
 TITLE OF INVENTION: HEMOPHILUS ADHESION PROTEINS
 NUMBER OF SEQUENCES: 6
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
 STREET: Four Embarcadero Center, Suite 3400
 CITY: San Francisco
 STATE: California
 COUNTRY: United States
 ZIP: 94111-4187
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/685,467
 FILING DATE: 22-JUL-1996
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/409,995
 FILING DATE: 24-MAR-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Silva, Robin M.
 REGISTRATION NUMBER: 38,304
 REFERENCE/DOCKET NUMBER: A-61053-2/RFT/RMS
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 781-1899
 TELEFAX: (415) 398-3249
 TELE: 910 277299
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1912 amino acids
 TYPE: amino acid
 STRANDEDNESS: unknown
 TOPOLOGY: unknown
 MOLECULE TYPE: protein

Query Match 5.0%; Score 94.5; DB 3; Length 1912;

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 11:20:48 ; Search time 20.94 Seconds
(Without alignments)
1656.554 Million cell updates/sec

Title: US-09-839-894-10
Sequence: 1886
1 MNKILFTFTLFSSVLTFTFA.....EAGQYMGNNVTFTPSSTQL 361

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 segs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:1*
1: PIR1:1*
2: PIR2:1*
3: PIR3:1*
4: PIR4:1*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1558.5	82.6	360	2	D56617 cfaf protein precu
2	949.5	50.3	363	2	S49539 Cood protein precu
3	930.5	49.3	364	2	S57937 Cold protein precu
4	275.5	14.6	359	2	AE0541 Probable fimbrial
5	120	6.4	1335	2	G90875 Probable factor [1
6	120	6.4	2660	2	E85822 Probable invas
7	114	6.0	650	2	S22835 alpha-eggulatin
8	109	5.8	2383	2	D64962 Probable membrane
9	108	5.7	524	2	S55097 Probable membrane
10	108	5.7	2529	2	B64635 toxin-like outer m
11	107	5.7	304	2	A64904 Probable fimbrial
12	107	5.7	304	2	C90892 Probable adhesin
13	107	5.7	304	2	F85725 Probable adhesin
14	106	5.6	671	2	A38109 auto-lysin - Entero
15	105.5	5.6	843	2	AC2507 hypothetical prote
16	105	5.6	598	2	AB1236 internatin protein
17	104.5	5.5	1441	2	B86807 hypothetical prote
18	103.5	5.5	691	2	B75622 hypothetical prote
19	103.5	5.5	1269	2	A90267 proteinase related
20	103.5	5.5	2399	2	H71879 toxin-like outer m
21	102	5.4	692	2	G90038 hypothetical prote
22	101	5.4	1238	2	AH0038 Probable exported
23	100	5.3	1752	2	T48965 hypothetical prote
24	99.5	5.3	3627	2	C83339 hypothetical prote
25	99	5.2	802	2	AH1580 phenylalanyl-tRNA
26	99	5.2	856	2	B81399 Probable periplasm
27	99	5.2	1849	2	C41859 Iga-specific metal
28	98.5	5.2	608	2	H64473 hypothetical prote
29	98.5	5.2	4919	2	T31105 hypothetical prote

30	98	5.2	587	2	AC1510 internatin protein
31	98	5.2	758	2	S47866 SP71 protein - ye
32	98	5.2	1014	2	S37405 cytotoxic necrotiz
33	98	5.2	1157	2	S38160 NUP133 protein - Y
34	97.5	5.2	1797	2	F69195 cell surface glyco
35	97	5.1	454	2	T01337 hypothetical prote
36	96.5	5.1	2364	2	I40884 cytotoxin L - Clos
37	96	5.1	992	2	T28421 Probable DNA-direc
38	95.5	5.1	713	1	ALBSXR cyclomaltodextrin
39	95.5	5.1	895	2	T02597 Mutator-like trans
40	95	5.0	413	1	S28602 translation releas
41	95	5.0	561	2	A84113 transposase (12) B
42	95	5.0	2334	2	S32920 cell wall-associat
43	95	5.0	5188	2	B85547 Probable RTX fami
44	94.5	5.0	835	2	E71691 outer membrane ass
45	94.5	5.0	858	2	T47223 replication licens

ALIGNMENTS

Query Match	82.6%	Score 1558.5	DB 2	Length 360
Best Local Similarity	81.2%	Pred. No. 1.2e-109		
Matches 293	Conservative 31	Mismatches 36	Indels 1	Gaps 1
QY	1	MNKLFTFTLFSSVLTFTFAVSADKIPGDSITNMGPRDRNBSKHNILNNHITAYSE 60		
DB	1	MNKLFTFTLFSSGFETFAVSADKNPGENMNTTIGPHRGSSPYILNLSYLVAYNG 60		
QY	61	SHLLYDMFTCLSSNHTLNGACPTSENPSSVSGFTNTLTPTFKRSILKRELQIKGY 120		
DB	61	SHLLYDMFTCLSSQNTLNGACPSDAPGTATIDGETNTLTPTFKRSILKRELQIKGY 120		
QY	121	KOLLFSVNCPSGLTNSAHFNCNKNAASGLYLYIPAGELNLPFGIMDTLKLRYK 180		
DB	121	KOLLFSVNCPSGLTNSAHFNCNKNAASGLYLYIPAGELNLPFGIMDTLKLRYK 180		
QY	181	RRSEYTGTYTNTTKLTKGNIQIWLPOFKSDARVDLRLPTGGTTYIGRNSVDMCFY 240		
DB	181	RRDYTYGTYTNTTKLTKGNIQIWLPOFKSNARVDLRLPTGGTTYIGRNSVDMCFY 240		
QY	241	DGYSTNSSLEIRPODNPKSDGKFYLRKINDPTKEIAYLTLGLGKSLTPNNGTSLI 300		
DB	241	DGYSTNSSLEIRPODNPKSDGKFYLRKINDPTKEIAYLTLGLGKSLTPNNGTSLI 300		
QY	301	ADAASETNMNRITTAATVMPETISVPVLCWPGRLDLDKAVENPEAGQYMGNNVTFTPSSTQL 360		
DB	301	ADAASETNMNRITTAATVMPETISVPVLCWPGRLDLDKAVENPEAGQYMGNNVTFTPSSTQL 360		
QY	361	L 361		

Db 360 L 360

RESULT 2

S49539

C:CoD protein precursor - Escherichia coli

C:Species: Escherichia coli

C:Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 10-Dec-1999

C:Accession: S49539

R:Prohlich, B.J.; Karakashian, A.; Meisen, L.R.; Wakefield, J.C.; Scott, J.R.

Mol. Microbiol. 12, 387-401, 1994

A:Title: CoD and CoD are required for assembly of CSI pill.

A:Reference number: S49538; MUID:94344028

A:Accession: S49539

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-363 <PRO>

A:Cross-References: EMBL:X76908; NID:9488735; PIDN:CAA54230.1; PID:9488737

C:Superfamily: Escherichia colonizing factor antigen cfaE

Query Match 50.3%; Score 949.5; DB 2; Length 363;

Best Local Similarity 53.1%; Pred. No. 7e-64; Mismatches 98; Indels 21; Gaps 11;

Matches 197; Conservative 55; Mismatches 98; Indels 21; Gaps 11;

QY 3 KILFIPLFFSSVLETPAVSADKIPGDE--SITNIF-GPR-DRNESSPKHNLNHNITAY 58

Db 2 KRIFIL-----SIFSAVVSAGRYPETVGNLTKSFQARLRDSVQSPITNITNIVAGY 57

QY 59 SESHTLYDKMTFLCLSSHNTLNGACPTSENSSSVS-GETNITLOFTEKRSILKRELQI 117

Db 58 SLSHSLYDRIVEFLCTSSSNPNVNCAPTI--GTSQVQYGTITITLOFTEKRSILKHNINL 114

QY 118 KGYKOLFFKSVNCPG--LTINSAHFNCKMNA-ASGASLYLYPAGELKLPFGIGMDAT 174

Db 115 AGKKRIWEMQSCDFNSNLNKSWSGAGHNGANTLNLTPAGEINKLPFGIGMEAT 174

QY 175 LKLRVRRYSET---YGYTITNITIKLTDKGNIOIWLPOFKSDAVDLNLRPTGGGTAY 230

Db 175 LILRLS-RGGEVSTHYGNVTNITVDLTDKGNIOIWLPOFHNPNVDLNLRIIGNWKYS 233

QY 231 GRSVVMCTYDGYSTNSSLIEIRFQNNRPKSDGKFLRKINDTKETIAYTLISLLAGKSL 290

Db 234 GSNLSLMCEYDGYSTNSDSVKIFQDNPNTNSEYMLYKKG-GTEKLPYAVSALMGEKIF 292

QY 291 TPNGTSLNIDAAASLETNMNRITAVTMPETISVPLCMPGRLOLDKVENPEAGQMGNT 350

Db 293 YPNQGSFTINDSSVLETNMNRITAVTMPETISVPLCMPGRLOLDKVENPEAGQMGNT 352

QY 351 NVYTFPSSQTL 361

Db 353 YTFPSEVNL 363

RESULT 3

S57937

C:CoD protein precursor - Escherichia coli

C:Species: Escherichia coli

C:Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 10-Dec-1999

C:Accession: S57937

R:Prohlich, B.J.; Karakashian, A.; Meisen, L.R.; Scott, J.R.

Submitted to the EMBL Data Library, January 1995

A:Description: The genes for CS2 pill of enterotoxigenic Escherichia coli and their inte

A:Reference number: S57934

A:Accession: S57937

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-364 <PRO>

A:Cross-References: EMBL:Z47800; NID:9897725; PIDN:CAA87763.1; PID:9897729

C:Genetics:

A:Start codon: TTG

C:Superfamily: Escherichia colonizing factor antigen cfaE

Query Match 49.3%; Score 930.5; DB 2; Length 364;

Best Local Similarity 48.4%; Pred. No. 1.9e-62;

Matches 177; Conservative 64; Mismatches 118; Indels 7; Gaps 3;

QY 1 MNKILFIPLFFSSVLETPAVSADKIPGDESITNIFGPRDRN--ESSPKHNLNHNITAY 58

Db 1 MKKIVFLVNLFCISQYGGSWHTNVAGSINKTESIGPIDRMAASYPAHYIFHEHVAGY 60

QY 59 SESHTLYDKMTFLCLSSHNTLNGACPTSENSSSVSGETNITLOFTEKRSILKRELQI 118

Db 61 NMDHSLFDRMTFLCLSSHNTLNGACPTSENSSSVSGETNITLOFTEKRSILKRELQI 118

QY 119 GYKOLFESKSVNCPG--PSGLTNSAHFNCKMNAASGASLYLYPAGELKLPFGIGMDAT 175

Db 119 GYKRFLESDDRCIHYDKNKLNSHTYKCGSFTRGVDFLLYIDGIDGLTGIGMEATL 178

QY 176 KLRVRRYSETGYTNTITIKLTDKGNIOIWLPOFKSDARVDLNRPTGGGTAYIGRNSV 235

Db 179 ELKVRKHVYNNHGYKVNTITVDLTDKGNIOIWLPTKFPSPRIDLNLRPENGNKYSQSNV 238

QY 236 DMCFTYDGYSTNSSLIEIRFQNNRPKSDGKFLRKINDTKETIAYTLISLLAGKSLPTNG 295

Db 239 EMLYDGYSTHQSISMRQDOSQGTGNNENLKTGEPLKPKYKLSLLGGEFFYNNNG 298

QY 296 TSLNIDAAASLETNMNRITAVTMPETISVPLCMPGRLOLDKVENPEAGQMGNTVFT 355

Db 299 EATITNDTSSLINMNRITAVTMPETISVPLCMPGRLOLDKVENPEAGQMGNTVFT 358

QY 356 PSSQTL 361

Db 359 PSSSSL 364

RESULT 4

AE0541

probable fimbrial protein tcfD [imported] - Salmonella enterica subsp. enterica serov

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Note: this species has also been called Salmonella typhi

C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001

C:Accession: AE0541

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church

th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr

, S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se

A:Reference number: AB0502; PMID:11677608

A:Accession: AE0541

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-359 <PAR>

A:Cross-References: GB:AL513382; PIDN:CAD08773.1; PID:g16501589; GSPDB:GN00176

C:Genetics:

A:Gene: tcfD

Query Match 14.6%; Score 275.5; DB 2; Length 359;

Best Local Similarity 29.7%; Pred. No. 3.1e-13;

Matches 90; Conservative 48; Mismatches 130; Indels 35; Gaps 11;

QY 70 FCLSSHNTLNGACPTSENSSSVSGETNITLOFTEKRSILKRELQIYKQLLFKSVN 129

Db 79 WVCRSNRNMEGACETHLVWYFAVYSKIRLRFPEDQSHAEITL-----ILGSVR 131

QY 130 --CPSGTLNSAHFNCKMNAAS--GASLYLYPAGELKLPFGIGMDATLKL-RYKRY 183

Db 132 DACYTG-----INMAAACQWGRSLKLRIPSEELAKIPISGWMKATVLVDLYQWGG 183

QY 184 SETGYTNTITIKLTD--KGNIOIWLPOF-KSDAVDLNLRPTGGGTAYIGRNSVDMCFY 240

Db 184 DDELGTSTIDITLNTVTDHAEANAATFFPGATRPATVVDLNLHRMNASQSGRANLDMCLY 243

```
QY 241 DGYSTNSSLEIFQDNNPKSDGFLYRKINDDTKEIAYTSLLAGSLPTNGSTLNI 300
  || : ||| : : || : : || : : || : : || : : || : : || : : || : :
Db 244 DG-GVAKRSLOMKIEGSKNSGTFQYIKSDADT--IDVAVSNMNGRSIPYTRGVEFSL 300
C:Accession: G90975
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Species: Escherichia coli
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasaara, N.; Yasunaga, T.; Kuhara, S.; Shida, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: G90975
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1335 <HAY>
A:Cross-References: GB:BA000007; PIDN:BA036198.1; PID:g13362243; GSPDB:GNO0154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: Ecs2775
```

RESULT 5

```
G90975
Probable factor [imported] - Escherichia coli (strain O157:H7, substrain RIMD 0509952)
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: G90975
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasaara, N.; Yasunaga, T.; Kuhara, S.; Shida, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: G90975
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1335 <HAY>
A:Cross-References: GB:BA000007; PIDN:BA036198.1; PID:g13362243; GSPDB:GNO0154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: Ecs2775

Query Match      6.4%; Score 120; DB 2; Length 1335;
Best local similarity 20.4%; Pred. No. 0.85;
Matches 80; Conservative 66; Mismatches 156; Indels 90; Gaps 15;
```

```
QY 13 SSYLTFPAVSADKIPGDESTITIFPRDRNESSPKHILNNHITAY---SESHITLYDRMT 69
  || : ||| : : || : : || : : || : : || : : || : : || : : || : :
Db 746 SAKIATLSASNNGVLANENANATVSVNVADEGS---NPINDHTVTFRAVLGSAISFNNO 802
C:Accession: G90975
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Species: Escherichia coli
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasaara, N.; Yasunaga, T.; Kuhara, S.; Shida, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: G90975
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1335 <HAY>
A:Cross-References: GB:BA000007; PIDN:BA036198.1; PID:g13362243; GSPDB:GNO0154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: Ecs2775

Query Match      6.4%; Score 120; DB 2; Length 1335;
Best local similarity 20.4%; Pred. No. 0.85;
Matches 80; Conservative 66; Mismatches 156; Indels 90; Gaps 15;
```

```
QY 13 SSYLTFPAVSADKIPGDESTITIFPRDRNESSPKHILNNHITAY---SESHITLYDRMT 69
  || : ||| : : || : : || : : || : : || : : || : : || : : || : :
Db 746 SAKIATLSASNNGVLANENANATVSVNVADEGS---NPINDHTVTFRAVLGSAISFNNO 802
C:Accession: G90975
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Species: Escherichia coli
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasaara, N.; Yasunaga, T.; Kuhara, S.; Shida, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: G90975
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1335 <HAY>
A:Cross-References: GB:BA000007; PIDN:BA036198.1; PID:g13362243; GSPDB:GNO0154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: Ecs2775
```

```
C:Accession: E85822
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
Hiller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; DiMantola, E.; Potamouzis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E85822
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2660 <STO>
A:Cross-References: GB:AE005174; NID:g12516151; PIDN:AA057041.1; GSPDB:GNO0145; UWGP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z3135
```

```
Query Match      6.4%; Score 120; DB 2; Length 2660;
Best local similarity 20.4%; Pred. No. 2.1;
Matches 80; Conservative 66; Mismatches 156; Indels 90; Gaps 15;
```

```
QY 13 SSYLTFPAVSADKIPGDESTITIFPRDRNESSPKHILNNHITAY---SESHITLYDRMT 69
  || : ||| : : || : : || : : || : : || : : || : : || : : || : :
Db 737 SAKIATLSASNNGVLANENANATVSVNVADEGS---NPINDHTVTFRAVLGSAISFNNO 793
C:Accession: G90975
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Species: Escherichia coli
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasaara, N.; Yasunaga, T.; Kuhara, S.; Shida, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: G90975
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1335 <HAY>
A:Cross-References: GB:BA000007; PIDN:BA036198.1; PID:g13362243; GSPDB:GNO0154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: Ecs2775

Query Match      6.4%; Score 120; DB 2; Length 2660;
Best local similarity 20.4%; Pred. No. 2.1;
Matches 80; Conservative 66; Mismatches 156; Indels 90; Gaps 15;
```

```
QY 13 SSYLTFPAVSADKIPGDESTITIFPRDRNESSPKHILNNHITAY---SESHITLYDRMT 69
  || : ||| : : || : : || : : || : : || : : || : : || : : || : :
Db 737 SAKIATLSASNNGVLANENANATVSVNVADEGS---NPINDHTVTFRAVLGSAISFNNO 793
C:Accession: G90975
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Species: Escherichia coli
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasaara, N.; Yasunaga, T.; Kuhara, S.; Shida, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: G90975
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1335 <HAY>
A:Cross-References: GB:BA000007; PIDN:BA036198.1; PID:g13362243; GSPDB:GNO0154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: Ecs2775

Query Match      6.4%; Score 120; DB 2; Length 2660;
Best local similarity 20.4%; Pred. No. 2.1;
Matches 80; Conservative 66; Mismatches 156; Indels 90; Gaps 15;
```


QY 174 TLKRVKRRSEYGYTITITIKLFDKGNIOIWLPOFKSD-ARVDLNLK-----PTGGCT 228
 Db 314 -----VETTEANNTYGL---VKLDDSGSL-----TIKDFEYNLESQNLNVLPTTKES 339
 QY 229 YIGRNSVDMC-----FYDGYSTNSSSL-----TRFODNNPKSDGK-----264
 Db 360 EISDSIYKCDNSAINTIYSGFTNNFTLPSQPAEIAIMLEYGVNGTNT-GKILLTDYAVP 418
 QY 265 ----FIRKINDTKELAYTLLSLLAGKSLTPNGTSLNTADAASLETNNR-ITAYTMP 319
 Db 419 TPENYTIKNNKDDT--ISATISYDKA-NSLNEIDVTATYVAKSASTSSQSSRSLSSTSP 475
 QY 320 EISVPILCMGRLQLDKAVENPEAGGYMGN-IVTFPPS 357
 Db 476 SSSTGSSSTG-----SSASSSSKSGKGVNIVVSEFSQS 510

RESULT 10

B64635
 toxin-like outer membrane protein HP0922 - Helicobacter pylori (strain 26695)
 C:Species: Helicobacter pylori
 C:date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
 C:Accession: B64635
 R:Tomb, J.F.; White, O.; Kervatage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.; Nature 388, 539-547, 1997
 A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
 A:title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
 A:Reference number: A64520; MUID:97394467
 A:Accession: B64635
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-2529 <TOM>
 A:Cross-references: GB:AE000602; GB:AE000511; NID:g2314060; PIDN:AAD07969.1; PID:g231406

Query Match 5.7%; Score 108; DB 2; Length 2529;

Best Local Similarity 22.5%; Pred. No. 16; Matches 84; Conservative 50; Mismatches 122; Indels 118; Gaps 22;

QY 85 TSENPSVSVSGETITLQTEKRSLLI-----KRELOIKGYKQLLFK---SVNCP 131
 Db 977 TSNFNATITQLGNTNFTL---SSQSLNFGDITLQNNANITIGNKSQAFAKNSLTIDNN 1033
 QY 132 SGLTL-NSAHFNCNKNA--SGASLYLYTPA-GELKNLPEGGIMDATLKLVRKRYSEY 187
 Db 1034 SNLSLDNOSVLANMNTSAFNQASLNIYNSQATFNSLPFNG---GTLNLASSKLNASN 1090
 QY 188 GTTITITITKLT-----KGNITQIWLPOFKSDARVDLNLKPTGGCTYIGRNSVD 236
 Db 1091 ASFSNNTTITLNDOSVLSASNTSSLANNI-----NFGQASQADP---GNTITIDTASFN 1140
 QY 237 MCFYDGYSTNSSSL-----EIRFODNNPK-----SDGKEYFRKIN---DDT 274
 Db 1141 --PDSASSLNFNLTANGALNFGYTPSLTKALMSVSGCVLGNNGDINLSIDNITDNT 1198
 QY 275 KEIAYTLLSLLAGKSLTPNGT-----LNIADAASLETN---NMRITAVTMPEI 321
 Db 1199 KSVTY--NILNAQGITGTSANGYEKILFYGKIKQIATYSDNNNTQTGTFINPLNSSQI 1256
 QY 322 -----SVPYL-----CMFGRQLDKAVENP-----EAGOT--M 347
 Db 1257 IOESIKNGDLTTEVLNPNNSASNTIFNIAPELYNQASKONPTGYSYSDNQAAGTYLT 1316
 QY 348 GNINVTFTP-SSOT 360
 Db 1317 SNIKGLTFPKGSOT 1330

RESULT 11

A64904
 probable fimbrial protein b1502 - Escherichia coli

C:Species: Escherichia coli

C:date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 29-Sep-1999

C:Accession: A64904

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: A64904

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-304 <BLAT>

A:Cross-references: GB:AE000247; GB:U00096; NID:g1787773; PIDN:AACT4575.1; PID:g17877

A:Experimental source: strain K-12, substrain MG1655

C:Superfamily: fimbrial protein fimH

C:Keywords: fimbria

Query Match 5.7%; Score 107; DB 2; Length 304;

Best Local Similarity 21.0%; Pred. No. 1.2; Matches 75; Conservative 51; Mismatches 125; Indels 106; Gaps 19;

QY 3 KILF-IETTFEVSLLFTFAVSADKIPGDESI---TNIEGPRDRNESSPKHNI---LNMH 54
 Db 8 KVLFGIYLLMAKRVAFSCNVN---GSSIGAGTTSVYNLD-PVIGPQGNLVLDLSCH 63
 QY 55 ITAVSESHLYDRMTFLCLSSH-NLNGACPTSENPSSSVSGETNITLOFTEKRSLIK 113
 Db 64 ISCMNDYGGWYD-----TDHINLVQG-----SAFAG-----89
 QY 114 ELQIKGYK-QLLKSVNCPGCLTLNSAHFNCNKNAASGASLYLYIRAGELKNLPEGGIMD 172
 Db 90 --SLQSTKGSLYNNWNTYPPPLTTNTNVDIGKTPMPLKLYI-----TPVGAAG 140
 QY 173 ATLK-----LRVRRYSEYGT-----YTITITIK-----LTKGNIOIWLPOFK 212
 Db 141 VVIKAGEVIAIHMYKATIGSGNPRNFTNIIISNNNVVPTGGCTYDSNRYVVDLPDF 200
 QY 213 SDARVDLNLKPTGGCTYIGRNSVDMCFYDGYSTNSSLETIRFODNNPKSDGKEYLRKIND 272
 Db 201 GSAEIPF-----GVYCS--SEOKLSEFLSGATVDSRQV-FANTAP-----D 239
 QY 273 DTEKIAVITISLLAGK-----SLTPNGTSLNTADAASLETNNRITAVTMPEI 321
 Db 240 ATRASGVVTLMRKILATGENVSLGTYNKSKVPJGLSATYQGTGKNYSAGTVOVS 296

RESULT 12

probable adhesin [imported] - Escherichia coli (strain 0157:H7, substrain RIMD 050995

C:Species: Escherichia coli

C:date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001

C:Accession: C90892

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.

gasaawara, N.; Yasunaga, T.; Kihara, S.; Shiba, T.; Hattori, M.; Shingawa, H.

DNA Res. 8, 11-22, 2001

A:title: Complete genome sequence of enterohemorrhagic Escherichia coli 0157:H7 and g

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: C90892

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-304 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA035530.1; PID:g13361573; GSPDB:GN00154

A:Experimental source: strain 0157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: ECS2107

C:Superfamily: fimbrial protein fimH

Query Match 5.7%; Score 107; DB 2; Length 304;

Best Local Similarity 21.0%; Pred. No. 1.2; Matches 75; Conservative 51; Mismatches 125; Indels 106; Gaps 19;

```

Oy 3 KILF-IFLFFSVLEFPAVSADKIPGDESI-----TNIJGPRDRNESSPKHNI-----LNNH 54
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 8 KVLRFITLLLAAGKVFASFSCVD---GSSSIGAGTTSYVNL-D-PIQPGQVLVVLSQH 63
Oy 55 ITVAESHTLTXDRMFLCLSSH-NTLNGACPTSENPSSSVSGEFTNITLOFETKRSLIKR 113
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 64 ISCMNDYGGWLD-----TDHINLYQG-----SARAG----- 89
Oy 114 ELQIKGYR-QLLEKSVNCPGSLTTLNSAHFNCNNKAASASLYLYIPAGELKRLNLPFGIMD 172
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 90 --SLQSGSLYXWNVVTFPLTNTNVLIDIGDKTPMLPLKLYI-----TPVGAAGG 140
Oy 173 ATLK-----LVRKRYSETYGH-----YTINITK-----LTDKNIQTLWPOFK 212
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 141 VYIKAGEVIARIHMXYKITLQSGNPRNFTWNIISNNSVYMPGCGTYPDSRNVTLNLPDF 200
Oy 213 SDARVDLMLRPTGGTLYIGRNSVDKCFYDGYSTNSSSLIEIRFQDNRPKSDGKFYLKRLND 272
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 201 GSAELPL-----GVYCS-SEQKLSFYLSGTTTDSARQV-FANTAP-----D 239
Oy 273 DTKELATYTLSTLLACK-----SLTPTNGSLNLIADAASLETWNRILTAWTMPEI 321
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 240 ATKASGVGVSILMRNCKIILATGENVSLGTVYKSKYPLGLSAYTGQGTGNKVSAGTQVOSV 296

```

RESULT 13
F85725
probable adhesin, fimb type protein Z2206 [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: F85725
R:Ferna, N.I.; Pinkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, U.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoultis, K.; Apodaca
Native 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: F85725
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-304 <STO>
A:Cross-references: GB:AE005174; NTD:G12515169; PION:AMG56266.1; GSFDB:GN00145; UWGP:Z22206
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z2206
C:Superfamily: fimbrial protein fimb

Query Match	5.7%	Score 107;	DB 2;	Length 304;
Best Local Similarity	21.0%	Pred. No. 1.2;		
Matches 75; Conservative	51;	Mismatches 125;	Indels 106;	Gaps 19

[illegible]

```

DB      24.0  ATKAGVGATLMRNKILATGENVSLGTIVNSKVPYELGATFATGQIGNKVMSAGTVQSV 296

RESULT  14

A38109      autoLysin - Enterococcus faecalis
C:Species:  Enterococcus faecalis
C:Date:     24-Jul-1992 #sequence_revision 24-Jul-1992 #text_change 15-Oct-1999
C:Accession: A38109
R:Beliveau, C.; Potvin, C.; Trudel, J.; Asselin, A.; Bellemare, G.
J. Bacteriol. 173, 5619-5623, 1991
A>Title:    Cloning, sequencing, and expression in Escherichia coli of a Streptococcus f
A:Reference number: A38109; MUID:91358349
A:Accession: A38109
A>Status:    preliminary
A:Molecule type: DNA
A:Residues: 1-671 <BR>
A:Cross-references: GB:580002; NID:9153658; PIDN:AAA67325.1; PID:9829194

```

[illegible]

hypothetical protein al17235 [imported] - Anabaena sp. (strain PCC 7120) plasmid pcc7120_02507
C:Species: Anabaena sp.
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:date: 14-Dec-2001 #sequence_rev1sion 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AC2507
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Matanabe, A.; Iriyama, N.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A:Reference numbers: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC2507
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-843 <KUU>
A:Cross-references: GB:BA000020; PIDN:BA878319.; PID:G17135773; GSPDB:GN00180
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: al17235
A:Genome: plasmid
Query Match 5.6% Score 105.5; DB 2; Length 843;

Best Local Similarity 18.9%; Pred. No. 5.8;
Matches 71; Conservative 67; Mismatches 127; Indels 111; Gaps 18;

```

QY 77 NTINGACPTSEN-----PSSSVSGETNITLQ-----FT--EKRSLIKREL 115
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 297 NILDGSLVFTQNHGFKTGAVKIDAOISNLAIQSSNLALSAIYISNFGFTPEGESIQLDVKDY 356
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 116 QIKGYQQLFKSVNCPGSL-TLNSAHFN-----CNKNAASGASLYIIP-----AG 160
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 357 TIQGGQIATTTTTPNAPSLITINSNLKISGDTPTSYANPDGLGINTFSTISSGKGGLIAG 416
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 161 ELKNLPPG--GIWD-----ATLKLVRKRRYSETGYTITITIKLTDKGNIOIWLIP 209
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 417 KINNTIIGLDGVFNVAASGAGCNLFLELENLIKDGASLGSSSTIRSGGGGNV----- 471
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 210 QEKSDARVDLN-----LRPT-----GGCTYIGRNSVDMCFYDGYSTNSSSLE--- 251
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 472 -FIKSONIDISQOSALLRPSNITSTFGHNGCNIDINTLNLIISNGCISSTLSAGKA 530
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 252 -----IRPDNNPK--SDGKFYLRKINDPTKEIAYTLISLLAGKS----- 289
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 531 GNISINSSNSINVGVTNINSNSPSTINSNFL--VDPNIDQLKLYROPPLIGQAGNIFL 588
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 290 ----LPTNGTSLN-----IADAASLETNMNRITAVTPEISVPLCWP-GRQLDAKYE 339
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 589 NTDIINISNGCLINARNEGVDAGNIRISANTININSQGEVNAATTIGEGNIIILNSR--- 646
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 340 NPEAGQYMGINIVTFT 355
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 647 ----NLFNNSRITAT 658
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: July 1, 2002, 11:24:57
Job time: 249 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 11:23:44 ; Search time 13.41 Seconds
(without alignments)

1042.338 Million cell updates/sec

Title: US-09-839-894-10

Perfect score: 1886
Sequence: 1 MNKILFIFLTFSSVLTFFA.....EAGQYMGNINVTTPSSQTL 361

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SWISSprot_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1558.5	82.6	360	1	CPAE_ECOLI
2	114	6.0	650	1	P20840 escherichia
3	108	5.7	524	1	GAS3_YEAST
4	107	5.7	304	1	YDEQ_ECOLI
5	106	5.6	671	1	ALYS_ENTFA
6	99	5.2	1849	1	IG44_HAEIN
7	98.5	5.2	608	1	YD93_METUA
8	98	5.2	752	1	SP21_YEAST
9	98	5.2	1157	1	NI33_YEAST
10	95.5	5.1	713	1	CD62_PAEWA
11	95	5.0	413	1	RFIM_YEAST
12	95	5.0	2334	1	WAPA_BACSU
13	94.5	5.0	959	1	NI00_YEAST
14	94	5.0	471	1	LEU2_BUCRP
15	93.5	5.0	412	1	THRG_SHEEP
16	93	4.9	1167	1	CLAA_BACTU
17	92.5	4.9	320	1	CYF_CYACA
18	92.5	4.9	544	1	AGM1_CANAL
19	92.5	4.9	1061	1	OAR_MYXXA
20	92.5	4.9	2278	1	FABL_YEAST
21	91.5	4.9	869	1	CFPC_ECOLI
22	91.5	4.9	1783	1	Y468_MYCGE
23	90.5	4.8	309	1	ELTB_CLOPE
24	90.5	4.8	863	1	MCW4_XENLA
25	90.5	4.8	1019	1	ENRK_HUMAN
26	90.5	4.8	1328	1	HUS2_SCHPO
27	90.5	4.8	2193	1	POG6_HE71M
28	89.5	4.7	1113	1	WGA2_YEAST
29	89	4.7	465	1	ANT3_MOUSE
30	88.5	4.7	455	1	MURF_BUCAT
31	88	4.7	1394	1	HAP_HAEIN
32	87.5	4.6	468	1	VGLC_HSVB
33	87.5	4.6	792	1	UBPA_YEAST

34	87	4.6	711	1	ETP2_SPKA
35	87	4.6	1290	1	BXCL_CLOBO
36	86.5	4.6	411	1	THRG_BOVIN
37	86.5	4.6	1002	1	SPHR_AMEPV
38	86.5	4.6	1273	1	WEI1_YEAST
39	86.5	4.6	1540	1	RM1_DROME
40	86.5	4.6	1769	1	YK9_YEAST
41	86	4.6	368	1	YB11_YEAST
42	86	4.6	493	1	FLIC_SALPA
43	86	4.6	828	1	MRC_KLEPN
44	86	4.6	1211	1	BUN2_DROME
45	86	4.6	1304	1	CD45_HUMAN

ALIGNMENTS

RESULT	1	STANDARD	PRT	360 AA.
CPAE_ECOLI				
ID	CPAE_ECOLI			
AC	P25734			
DT	01-MAY-1992 (Rel. 22, Created)			
DF	01-MAY-1992 (Rel. 22, Last sequence update)			
DT	01-AUG-1992 (Rel. 23, Last annotation update)			
DE	CPA/I fimbrial subunit E (Colonization factor antigen I subunit E).			
GN	CPAE.			
OS	Escherichia coli.			
OG	Plasmid NTP513.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OX	NCBI_TaxID=562;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92329981; PubMed=1352712;			
RA	Jordi B.J.A.M., Willshaw G.A., van der Zeijst B.A.M., Gaastra W.;			
RT	"The complete nucleotide sequence of region I of the CPA/I fimbrial			
RL	operon of human enterotoxigenic Escherichia coli.";			
RM	Microb. Pathog. 6:297-309(1989).			
RN	(2)			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92329981; PubMed=1352712;			
RA	Jordi B.J.A.M., Willshaw G.A., van der Zeijst B.A.M., Gaastra W.;			
RT	"The complete nucleotide sequence of region I of the CPA/I fimbrial			
RL	operon of human enterotoxigenic Escherichia coli.";			
RM	DNA Seq. 2:257-263(1992).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation-			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: M55661; AAC41417.1; -			
KW	Antigen; Fimbria; Plasmid.			
SO	SEQUENCE 360 AA; 39903 MW; 69150963A8E69CE CRC64;			

Query Match 82.6%; Score 1558.5; DB 1; Length 360;
Best local similarity 81.2%; Pred. No. 1.4e-112;
Matches 233; Conservative 31; Mismatches 36; Indels 1; Gaps 1;

QY	1	MNKLIFLTFSSVLTFFAVSADKIPGDESTINIRPPDRNRPSPKHNLTNNHTAYSE 60
DB	1	MNKLIFLTFSSVLTFFAVSADKIPGDESTINIRPPDRNRPSPKHNLTNNHTAYSE 60
QY	61	SHLLYDRMPFLCLSSHTNLGACPTSENPSSVSGETNTTLQFTKRSLLIKRELQIKGY 120
DB	61	SHLLYDRMPSLCLSSQNTLNGACPSDAPETATIDGENTTLQFTKRSLLIKRELQIKGY 120

OY	121	KOLLEKSYNCSGGLTNLNAHFHNCNNKNAASGATLYTIPAGELKTLPGGIMDATLKLRYK	180
Db	121		180
OY	121	KQFLFKNANCSKTLANSHFQCNREQASGATLSLYIPAGELINKLPPGCGWNAVYLKLNK	180
OY	181	RRYSFTGYTYINTITIKLTDKGNIOIWLPOKSDARVDLNLPTGGGTYIGRNSVDMCFY	240
Db	181		240
OY	241	DGYSTNSSLEFRPDNNPKSDGKFYLKINDDKELVYTLISLLAKRSLTPPNNGISLNT	300
Db	241		300
OY	301	ADAALEETNMNRITAVTPELSIVLGCWPGRLQDAKVENPEAGQYNGINVTFTPESSQT	360
Db	301		360
OY	361	L 361	
Db	360	L 360	
RESULT	2		
SAGL_YEAST	STANDARD:	PRT:	650 AA.
ID	SAGL_YEAST		
AC	P20840:		
DT	01-FEB-1991 (Rel. 17, Created)		
DT	01-NOV-1995 (Rel. 32, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Alpha-agglutinin precursor (AG-alpha-1).		
CN	SAG1 OR AGA1 OR YTR004C OR J1418.		
OS	Saccharomyces cerevisiae (Baker's yeast).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
OX	NCB1_TaxID=4932;		
LN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=90014768; PubMed=2677666;		
RA	Lipke P.N., Wojciechowicz D., Kurjan J.;		
RT	"Ag alpha 1 is the structural gene for the Saccharomyces cerevisiae		
RT	alpha-agglutinin, a cell surface glycoprotein involved in cell-cell		
RT	interactions during mating.";		
RL	Mol. cell. Biol. 9:3155-3165(1989).		
LN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-S288C;		
RX	MEDLINE=90005993; PubMed=2676603;		
RA	Hauser K., Tanner W.;		
RT	"Purification of the inducible alpha-agglutinin of S. cerevisiae and		
RT	molecular cloning of the gene.";		
RL	Fems Lett. 255:290-294(1989).		
LN	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-S288C / FY1679;		
RA	de Haan M., Smits P.H.M., Grivell L.A.;		
RT	Submitted (MAY-1995) to the EMBL/Genbank/DBJ databases.		
LN	[4]		
RP	PARTIAL SEQUENCE, CARBOHYDRATE-LINKAGE SITES, AND DISULFIDE BONDS.		
RX	MEDLINE=9004684; PubMed=7592821;		
RA	Chen M.-H., Shen Z.-M., Bobin S., Kahn P.C., Lipke P.N.;		
RT	"Structure of Saccharomyces cerevisiae alpha-agglutinin. Evidence for		
RT	a yeast cell wall protein with multiple immunoglobulin-like domains		
RT	with atypical disulfides.";		
RL	J. Biol. Chem. 270:26168-26177(1995).		
CC	-1- FUNCTION: CELL SURFACE GLYCOPROTEIN PROMOTING CELL-CELL CONTACT		
CC	TO FACILITATE MATING. SACCAROMYCES CEREVISIAE A AND ALPHA CELLS		
CC	EXPRESS THE COMPLEMENTARY CELL SURFACE GLYCOPROTEINS A-AGGLUTININ		
CC	AND ALPHA-, RESPECTIVELY, WHICH INTERACT WITH ONE ANOTHER TO		
CC	PROMOTE CELLULAR AGGREGATION DURING MATING.		
CC	-1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR		
CC	(POSSIBLE).		
CC	-1- INDUCTION: BY EXPOSITION TO PHEROMONE (A-FACTOR) SECRETED BY THE		
CC	OPPOSITE MATING TYPE CELLS (TYPE A).		
CC	-1- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED.		

	-!	SIMILARITY:	TO C.ALBICANS ALST.
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see http://www.isb-sdb.ch/announce/		
CC	or send an email to license@sdb.slb.ch).		
CC	-----		
DR	EMBL; M28164; AAA34417.1; -;		
DR	EMBL; X16861; CAA34752.1; -;		
DR	EMBL; X87611; CAA60926.1; -;		
DR	EMBL; Z49504; CAA89526.1; -;		
DR	PIR; S22835; S22835.		
DR	SGD; S0003764; SRG1.		
KM	Glycoprotein; Cell adhesion; Signal; GPI-anchor; Repeat.		
FT	SIGNAL	1	19
FT	CHAIN	20	650
FT	DOMAIN	278	350
FT			
FT	DOMAIN	339	423
FT	REPEAT	339	378
FT	REPEAT	384	423
FT	DISULFID	97	114
FT	DISULFID	202	300
FT	CARBOHYD	79	79
FT	CARBOHYD	109	109
FT	CARBOHYD	135	135
FT	CARBOHYD	248	248
FT	CARBOHYD	282	282
FT	CARBOHYD	289	289
FT	CARBOHYD	299	299
FT	CARBOHYD	303	303
FT	CARBOHYD	306	306
FT	CARBOHYD	307	307
FT	CARBOHYD	308	308
FT	CARBOHYD	311	311
FT	CARBOHYD	314	314
FT	CARBOHYD	315	315
FT	CARBOHYD	316	316
FT	CARBOHYD	329	329
FT	CARBOHYD	331	331
FT	CARBOHYD	334	334
FT	CARBOHYD	335	335
FT	CARBOHYD	338	338
FT	CARBOHYD	339	339
FT	CARBOHYD	340	340
FT	CARBOHYD	341	341
FT	CARBOHYD	342	342
FT	CARBOHYD	345	345
FT	CARBOHYD	346	346
FT	CARBOHYD	349	349
FT	CARBOHYD	350	350
FT	CARBOHYD	364	364
FT	CARBOHYD	402	402
FT	CARBOHYD	460	460
FT	CARBOHYD	485	485
FT	CARBOHYD	501	501
FT	CARBOHYD	614	614
FT	CONFLICT	449	449
FT	CONFLICT	556	556
FT	CONFLICT	581	581
SO	SEQUENCE	650 AA;	70339 MW; 8BBF7A1C4C93C2B CRC64;
	Query Match	Best Local Similarity	6.0%; Score 114; DB 1; Length 650;
	Matches	98; Conservative	Pred. No. 0.24; Mismatches 147; Indels 172; Gaps 23;
Oy	4	I L F I T P L F S S V L -----	T T F A ----- V S A D K I P G ----- D E S I T I N I F G R D R N E --- 43
		: : :	: : :
D	B	8	I L F I T P L A S I A N I N D T F S L U L E T P L A N O P O G W T A F D A S S I R E G D E F T L 67

FT	SIGNAL	1	21	POTENTIAL.
FT	CHAIN	22	?	GAS3 PROTEIN.
FT	PROPEP	?	524	REMOVED IN MATURE FORM.
FT	DONAIN	458	496	SER-RICH.
FT	CARBOHYD	201	201	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	269	269	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	350	350	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	385	385	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	404	404	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	422	422	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	504	504	N-LINKED (GLCNAC. . .) (POTENTIAL).
SO	SEQUENCE	524 AA;	56793 MW;	3A2BED0BD3ED8690 CRC64;
Query Match				
Best local Similarity		5.7%;	Score 108;	DB 1; Length 524;
Matches 69;		Conservative 42;	Mismatches 96;	Indels 72; Gaps 18;
QY	119 GYKQL-LFKSVNCSSGLTNSAHNCKKMA---SGASLYIYIPAGELKNLPFGGINDA	173		
Db	264 GYDKLNSTFEDAVIP---LIFSEGCKNKPRTPEDEVESEGLY---GGIKNFSSGL---	313		
QY	174 TLKLKVRKRSYETGYTIINTIKTLTDKNGIOIMLPKPRSD-ARYDLNLR---PTGGGT	228		
Db	314 -----VEYIEEANNKGL---VKLDDSGSL-----TYKDDPVNLIESQLKNVSLPTKES	359		
QY	229 YIGRNSVDMC-----FYDGYSTNSSSL-----IRPDNNPKRSDGK-----	264		
Db	360 EISDSIYKCDMSATNTIYSGFGTNNFLPSPAFIANNIEGVNGTWF-GKILDYAVP	418		
QY	265 ----FYLRKINDDKREIAYTISLLIAGKSLPFTNGTSLNIADASLETNMNR-TTAVTWP	319		
Db	419 TTFNTYTIKNNDDT--ISATISYDKA-NSLMLDVTATTVAAKSASTSOSSSRSLSSTSP	475		
QY	320 EISVPVLCMPGRLOLDAAKVENPEAGQVGNI-NVTFPTS	357		
Db	476 SSSSTGSSSTG---SSSASSSSKSGVGNVWSFSQS	510		
RESULT 4				
YDEQ_ECOLI	YDEQ_ECOLI	STANDARD;	PRT;	304 AA.
AC	P77588;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Hypothetical fibrial-like protein ydeq precursor.			
GN	YDEQ OR B1502.			
OS	Escherichia coli.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Escherichia.			
OX	NCBI_TaxID=562;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-K12 / MG1655;			
RX	MEDLINE=97251357; PubMed=9097039;			
RA	Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,			
RA	Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,			
RA	Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,			
RA	Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshida T., Saito N.,			
RA	Sampel G., Seki Y., Sivasubram S., Tagami H., Takeda J.,			
RA	Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horuchi T.,			
RT	"A 570-kb DNA sequence of the Escherichia coli K-12 genome			
RL	Science 277:1453-1474(1997).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-K12;			
RX	MEDLINE=97251357; PubMed=9097039;			
RA	Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,			
RA	Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,			
RA	Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,			
RA	Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshida T., Saito N.,			
RA	Sampel G., Seki Y., Sivasubram S., Tagami H., Takeda J.,			
RA	Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horuchi T.,			
RT	"A 570-kb DNA sequence of the Escherichia coli K-12 genome			

RT corresponding to the 28.0-40.1 min region on the linkage map.
 RL DNA Res. 3:363-377(1996).
 CC -1- SIMILARITY: TO TYPE-1 FIMBRIAL SUBUNITS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: AE000247; AAC74575.1;
 DR EMBL: D90792; BAA15175.1;
 DR EMBL: D90793; BAA15183.1;
 DR EcoGene: EG13799; YdeQ.
 DR InterPro: IPR000259; Fimbril.
 DR Pfam: PF000419; Fimbril; 1.
 DR Hypothetical protein; Fimbril; Signal; Complete proteome.
 FT SIGNAL 1 26
 FT CHAIN 27 304
 FT SEQUENCE 304 AA; 32069 MW; 8153C86E3087D99A CRC64;

Query Match 5.7%; Score 107; DB 1; Length 304;
 Best Local Similarity 21.0%; Pred. No. 0.32;
 Matches 75; Conservative 51; Mismatches 125; Indels 106; Gaps 19;

QY 3 KILF-IPTLFSSVLEFAVSADKIPGDESI---TNIIFGRPNESPKHNT---LNH 54
 DB 8 KVLFGITLLMAGKVFVFCNVND---GSSIGAGTTSVYVNL-DPVIOQONLVDSQH 63
 QY 55 ITAVSESHTLYDRMTFLCLSSH-NTLNGACPTSENPSSVSGETNITLQFTEKRLIKR 113
 DB 64 ISCMNDYGMWD-----TDHNLVYG-----SAFAG----- 89
 QY 114 ELQIKGK-QLLKSVCPCGGLTNSAHFNCNNKNAASGASLYLYIPAGELKNLPEFGIMD 172
 DB 90 -SLQSGKGLYNNVNYPPPLTNTNVDIGDKTPMPLKLYI-----TPVGAAG 140
 QY 173 ATLK-----LRVRRYSEYGT-----YTINITIK-----LTDKNIDIMLPQK 212
 DB 141 VVKAEVARIKIMKATIGSGNPRNFTWNITSNNNVMPGCGCYVDSNVVDLPDP 200
 QY 213 SDARVDLNLPTGGTYIGRNSVDMCFYDGYSTNSSLEIRFODNNPKSGKFLKIND 272
 DB 201 GSAEIPL-----GYVCS-SEQKLSPLYSGATTDSRQV-FANTAP-----D 239
 QY 273 DREIAYTSLSLAGK-----SLTPNGISLNTADAASLETNNRITAVTMEI 321
 DB 240 ATRKASGVTLMRNKILATGENVSLGTVNRKVPILGISTATYQGTGNKVSAGTVQSV 296

RESULT 5

ALYS_ENTRA STANDARD: PRT; 671 AA.
 AC P37710;
 DT 01-OCT-1994 (Rel. 30; Created)
 DT 01-OCT-1994 (Rel. 30; Last sequence update)
 DT 16-OCT-2001 (Rel. 40; Last annotation update)
 DE Autolysin precursor (EC 3.2.1.-) (peptidoglycan hydrolase)
 DE (Beta-glycosidase).
 OS Enterococcus faecalis (Streptococcus faecalis).
 CC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
 CC Enterococcus.
 OX NCBI_TaxID=1351;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91358349; PubMed=1679432;
 RA Bellevue C., Potvin C., Trudel J., Asselin A., Bellemare G.;
 RT "Cloning, sequencing, and expression in Escherichia coli of a
 RT Streptococcus faecalis autolysin."
 RL J. Bacteriol. 173:5619-5623(1991).

CC -1- FUNCTION: HYDROLYZES THE CELL WALL OF E.FAECALIS AND
 CC M.LYSODEIKTUS. MAY PLAY AN IMPORTANT ROLE IN CELL WALL GROWTH
 CC AND CELL SEPARATION.
 CC
 CC -1- SUBCELLULAR LOCATION: Secreted (Probable).
 CC
 CC -1- DOMAIN: LYSM REPEATS ARE THOUGHT TO BE INVOLVED IN PEPTIDOLYCAN
 CC BINDING.
 CC
 CC -1- SIMILARITY: BELONGS TO FAMILY 73 OF GLYCOSYL HYDROLASES.
 CC
 CC -1- SIMILARITY: CONTAINS 5 LYSM REPEATS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M58002; AAA67325.1;
 DR PIR: A38109; A38109.
 DR InterPro: IPR002901; Amidase_4.
 DR InterPro: IPR002482; LysM.
 DR Pfam: PF01476; LysM; 4.
 DR Pfam: PF01476; LysM; 4.
 DR SMART: SM00047; LysM; 5.
 DR SMART: SM00257; LysM; 5.
 DR SMART: SM00257; LysM; 5.
 DR Hydrolase; Glycosidase; Bacteriolytic enzyme; Cell wall;
 KW Cell division; Septation; Repeat; Signal.
 FT SIGNAL 1 53
 FT CHAIN 54 671
 FT REPEAT 363 405
 FT REPEAT 431 473
 FT REPEAT 499 541
 FT REPEAT 567 609
 FT REPEAT 629 671
 FT SEQUENCE 671 AA; 70442 MW; 3456268C6C1AA33 CRC64;

Query Match 5.6%; Score 106; DB 1; Length 671;
 Best Local Similarity 23.6%; Pred. No. 1;
 Matches 73; Conservative 39; Mismatches 103; Indels 94; Gaps 17;

QY 41 RNESPKHNT-LNNHITAVSESHTLYDRMTFLCLSSHNTLNGACPTSENPSSSV----- 94
 DB 309 RYATDPSYNAKLNNNVITAY--NLQYDTPSSGNTGGGVNPGTGSNNQSGNTYYTVK 366
 QY 95 SGFT--NITLQF-----TEKSL--IKRELQIKYKOLLKRSVNCPCGGLTNSAHFNCNNK 146
 DB 367 SGDTLKKIAAQGVSVANLRSMNGISGDLLEVQOKLIVKGA--SGNTGGSGNGGNNN 423
 QY 147 AASGASLYLYIPAGELKNLPEFGIMDATLKRVRYSFTGYTYTINITIKLTDKNIOI 206
 DB 424 -QSGTNTYYTVKSGDTLN-----KIAAQG-----VIV-----ANLNS 455
 QY 207 WLPOFKSDARVDLNL-----RPTGGGTYIGRNSVDMCFYDGYSTNSSLEIRFODNN 258
 DB 456 W-----NGISGDLIFVGOKLIVKKTSGNT-----GGSSNGS-----NNN 491
 QY 259 PKSDGKFLKINDDKRELA--YTLN-----LLAGKSLTPNGTSLIAIDA 303
 DB 492 QSGTNTYYTVKSGDTLNKIAAQGVSVANLRSMNGISGDLIFAGOKIIVKKTSGNTGGS 551
 QY 304 ASLETNNMR 312
 DB 552 SNGGSNNNQ 560

RESULT 6

IG44_HAEIN STANDARD: PRT; 1849 AA.
 AC P45386;
 DT 01-NOV-1995 (Rel. 32; Created)
 DT 01-NOV-1995 (Rel. 32; Last sequence update)
 DT 01-NOV-1995 (Rel. 32; Last annotation update)

[illegible]

OY 229 YIGNSVDMCFYDYSNSSLERFODNNKSGKRYLRKINDTKEIAYT-----LSL 283
 Db 259 YTNRTGTLMKY--GFAV-----IFFEENGTKSGTKI-----EGIATGYGVSA 300
 OY 284 LLAGKSLPTNG-----TSINADASLEENMKRITAYMPEI--SYPUICMGRIOL 334
 Db 301 TKESPELNASGKYEIWEYSANVSNKAS--SYEPNLTHVTIWAANGNPILLDEFNITL 357

RESULT 8
 SP21_YEAST STANDARD: PRT: 758 AA.

AC P35209;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE SPT21 protein.
 GN SPT21 OR YMR179W OR YMR010.09.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94186069; PubMed=8138180;
 RA Matsoulis G., Winston F., Boeke J.D.;
 RT "The SPT10 and SPT21 genes of Saccharomyces cerevisiae.";
 RL Genetics 136:93-105(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S28C / AB972;
 RA Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.

-1- FUNCTION: REQUIRED FOR NORMAL TRANSCRIPTION AT A NUMBER OF LOCI IN YEAST.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: L24436; AAA35078.1; -
 DR EMBL: Z49808; CAA89912.1; -
 DR PIR: S47866; S47866.
 DR SGD: S0004791; SPT21.
 FT DOMAIN 127 144 ASF/GLU-RICH (ACIDIC).
 FT DOMAIN 672 682 ASF/GLU-RICH (ACIDIC).
 FT SEQUENCE 758 AA; 84697 MW; 7DB3FCFE996705 CRC64;

Query Match 5.2%; Score 98; DB 1; Length 758;
 Best Local Similarity 21.9%; Pred. No. 5;
 Matches 61; Conservative 34; Mismatches 95; Indels 88; Gaps 12;

OY 42 NESSPKHNIINNHITAYSESHLYDRMTF-----LCSSHNTLNGACPSENSSSVS 95
 Db 166 NISNKKGRVNNNOI-----PETLEVKLRFTKYITNLTSNNNTNSRISCLQWPSL-- 218
 OY 96 GETNITLOFTEKRSLLKRELQIKGYKQLFKSVNCPGSLTNSAHFNCKNNAASGASLYL 155
 Db 219 --PSATLPLFPKQSLSFKTNQIK-----NSKNARTITIIIN-----NTMSGT-- 257
 OY 156 YIPAGELKNLPGGIDWATLKLAVKRRYSFTYTYTINITIKLDKGNIOIWLDPQEKSDA 215
 Db 258 --VGRROTNPM-----PAPKAVRTO-----SLPIW----- 280
 OY 216 RVDLNRPPGGGTYIGRNSVDMCFY-----DGYTNSSSLERFODNNKSGKRYLRK 269
 Db 281 ----NLKPNIANTGPPRNSIAHKIYLADRKTEANQONHONITAYEINTLQNDNTIORTK 336

OY 270 INDD-TKEIAYTLLSLAGSLPTNGT-----SLNI 300
 Db 337 IDDSVSKRPFDMKKRSKTKVSPGIATIKKPRASINI 374

RESULT 9
 N133_YEAST STANDARD: PRT: 1157 AA.

AC P36161;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Nucleoporin NUP133 (Nuclear pore protein NUP133).
 GN NUP133 OR YKR082W OR YKR402.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JUX1R;
 RX MEDLINE=95112817; PubMed=7813444;
 RA Doye V., Wept R., Hurt E.C.;
 RT "A novel nuclear pore protein Nup133p with distinct roles in poly(A)+
 RT RNA transport and nuclear pore distribution.";
 RL EMBO J. 13:6062-6075(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94262327; PubMed=8203164;
 RA Garcia-Cantalejo J., Baladron V., Esteban P.F., Santos M.A., Bou G.,
 RA Remacha M.A., Revuelta J.L., Ballesta J.P.G., Jimenez A., del Rey F.;
 RT "The complete sequence of an 18,002 bp segment of Saccharomyces
 RT cerevisiae chromosome XI contains the HBS1, MRP-120 and PRP16 genes,
 RT and six new open reading frames.";
 RL Yeast 10:231-245(1994).

-1- FUNCTION: INVOLVED IN POLY(A)+ RNA TRANSPORT AND NUCLEAR PORE DISTRIBUTION.
 -1- SUBCELLULAR LOCATION: Nuclear pore complex.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC

DR EMBL: X80066; CAA56372.1; -
 DR EMBL: Z27116; CAA81633.1; -
 DR EMBL: Z28307; CAA82161.1; -
 DR PIR: S38160; S38160.
 DR PIR: S39123; S39123.
 DR SGD: S0001790; NUP133.
 DR Nuclear protein; Transport; Transmembrane.
 KW Nuclear protein; Transport; Transmembrane.
 FT TRANSMEM 217 233 POTENTIAL.
 FT SIMILAR 98 420 TO NUP120 (AA 434-763).
 FT SEQUENCE 1157 AA; 133319 MW; C8BDBB7D709C5C08 CRC64;

Query Match 5.2%; Score 98; DB 1; Length 1157;
 Best Local Similarity 22.1%; Pred. No. 8.5;
 Matches 75; Conservative 43; Mismatches 127; Indels 94; Gaps 21;

OY 48 HNLIINNHITAY-----SESHLYDRMTF-----LCSSHNTLNGACPSEN 89
 Db 96 YGLVNHKKRYIWNHSTOKDPTIYIVPFRSDNDDEIAVAPRCILTFPATMDESPALNP 155
 OY 90 SSSVSAGE-----TNITLOFTEKRSLLKRELQIK-----GYKQLLRK 126
 Db 156 NDDDEFGGLIITIGSKAIYIEDINSLNMLFKLSERFS--HELELPINSSGGEKCDLM- 211
 OY 127 SYNC-PSGLTINS-----AHFNCKNNAASGASLYLYIPAGELKNLPGG-GIWDATTLKLR-- 178

Db 212 -LNCBPAGIVLSTNMGRIFETIRNSMGKPOLKLV---GKLNKPKFKLGIMSKIFENTNSS 266
 QY 179 -KKRRSEYV---GYTINTITIKLTKGNQIWLPOFKSDARDNLNRPFGGGTYIGRNAY 235
 Db 267 VVSLRNGPILGKGLRVLYTT---TNKGIFQTV--QLSA-----TNSHPV-----KLI 308
 QY 236 DMCIFYDGYSTNNSSTV-----EIRFODNPKPSD--GKFLRKINDPT-KEIAYTSLILLA 286
 Db 309 DVNIETALIESQDLYPRAHGLKLTWDSHPLODESSQFLSSITDSSCNEMTYIISTITIF 368
 QY 287 GKSLTPFNGTSLNIADAASLETNNRITAVTM-PEISVP 324
 Db 369 DSS-----SNSFTIFSTYRLNTFMESITDTKPKPIFIP 402
 RESULT 10
 CDG2_PAEMA STANDARD: PRT: 713 AA.
 ID CDG2_PAEMA STANDARD: PRT: 713 AA.
 AC P31835;
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 01-JUL-1993 (Rel. 37, Last annotation update)
 DE Cyclomalotodextrin glucanotransferase precursor (EC 2.4.1.19)
 DE (Cyclodextrin-glycosyltransferase) (CGTase).
 OS Paenibacillus macerans (Bacillus macerans).
 OC Bacteria: Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Paenibacillus.
 NCBI_TaxID=44252;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-37.
 RA Sugimoto T., Kubota M., Sakai S.;
 RT "Polypeptide possessing cyclomalotodextrin glucanotransferase activity.";
 RL Patent number GB2169902, 23-JUL-1986.
 CC -1- CATALYTIC ACTIVITY: Degrades starch to cyclodextrins by formation of a 1,4-alpha-D-glucosidic bond.
 CC -1- COFACTOR: BINDS TWO CALCIUM IONS.
 CC -1- SUBUNIT: MONOMER.
 CC -1- MISCELLANEOUS: CGTASE MAY CONSIST OF TWO PROTEIN DOMAINS: THE ONE IN THE AMINO-TERMINAL SIDE CLEAVES THE ALPHA-1,4-GLUCOSIDIC BOND IN STARCH, AND THE OTHER IN THE C-TERMINAL SIDE CATALYZES OTHER ACTIVITIES, INCLUDING THE RECONSTITUTION OF AN ALPHA-1,4-GLUCOSIDIC LINKAGE FOR CYCLIZING THE MALTOOLIGOSACCHARIDE PRODUCED.
 CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO KNOWN AS THE ALPHA-AMYLASE FAMILY.
 DR PIR: S26589; ALBSXR.
 DR HSSP: P43379; 1CDG.
 DR InterPro: IPR000461; Alpha-amylase.
 DR InterPro: IPR002044; CBD_4.
 DR InterPro: IPR002809; IPT_TIG.
 DR Pfam: PF00128; alpha-amylase; 1.
 DR Pfam: PF02806; alpha-amylase_C; 1.
 DR Pfam: PF01833; TIG; 1.
 DR Pfam: PD001568; CBD_4; 1.
 DR Pfam: PD001568; CBD_4; 1.
 KW Transferrase; Glycosyltransferase; Calcium; Signal.
 FT SIGNAL 1 27
 FT CHAIN 28 713 CYCLOMALOTODEXTRIN GLUCANOTRANSFERASE.
 FT DOMAIN 28 165 A1.
 FT DOMAIN 166 229 B.
 FT DOMAIN 230 434 A2.
 FT DOMAIN 435 522 C.
 FT DOMAIN 523 609 D.
 FT DOMAIN 610 713 E.
 FT ACT_SITE 256 256 BY SIMILARITY.
 FT ACT_SITE 285 285 BY SIMILARITY.
 FT ACT_SITE 356 356 BY SIMILARITY.
 SQ SEQUENCE 713 AA; 76857 MW; 5A287BC4AAFE635 CRC64;

Query Match 5.1%; Score 95.5; DB 1; Length 713;
 Best Local Similarity 20.1%; Pred. No. 7.2;

Matches 65; Conservative 43; Mismatches 131; Indels 85; Gaps 12;
 QY 68 KTFELSSNHTLNGACPTSENPSSSVSGETNTTLOFTFKRLKLELDIKGKOLLFS 127
 Db 8 LTVSVSSVGIAGALPVAASP-DTSVNNKLNFS--TDTVYQIVDRFVDG-----NS 57
 QY 128 VNCPSGLTNSAHFNCKNNAAGASLYLYIPAGELKNLPFGGIWDATLKVRRRYSEY 187
 Db 58 ANNPCTAAAFSSDHSN-----LKLTV-----FGGDMQG-----ITNKINDGY 92
 QY 188 GYTTINTITIKLTKDGNQIWLPOFKSDARDNLNRPFGGGTYIG----- 231
 Db 93 -----LTCMGITLALMISOPVENTAVIVSGVNNATAHYGWPDRFKKTAAGFSG 142
 QY 232 -----RNSVDMCFYDGYSTNNSSEIRFODNPKPSDGFYLRKINDPTKEA 278
 Db 143 TDFSNIILAAHSHNIRKVVDFAPNHNTPASSSTDPSFENGALYNCTLLGKYSNDPAGLE 202
 QY 279 YTLISLLAGKSLTPV-NGTSLNIADAASLETNNRITAVTMPEISVPLCMFGRLQLDK 337
 Db 203 HH-----NGSTDFTSTESGILYKMLYDLADINONNNTIDSTLKESTQLMLNLGVGDI 258
 QY 338 VENPENGQ--YMGNI-----NVTFT 355
 Db 259 KHPQGMQKSYVSSISYSANPVFT 282
 RESULT 11
 ID REIM_YEAST STANDARD: PRT: 413 AA.
 AC P30775;
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 01-JUL-1993 (Rel. 41, Last annotation update)
 DE Peptide chain release factor 1, mitochondrial precursor (MRF-1).
 DE MRF1 OR YGL143C.
 GN MRF1 OR YGL143C.
 OS Saccharomyces cerevisiae (Baker's Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-93117110; PubMed=1475194;
 RA Pel H.J., Maat M.J., Rep M., Grivell L.A.;
 RT "The yeast nuclear gene MRF1 encodes a mitochondrial peptide chain release factor and cures several mitochondrial RNA splicing defects.";
 RT Nucleic Acids Res. 20:6339-6346(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / FY1769;
 RX MEDLINE-97197983; PubMed=9046099;
 RA Voet M., Defoor E., Verhasselt P., Riles L., Robben J., Volckaert G.;
 RT "The sequence of a nearly unclonable 22.8 kb segment on the left arm of chromosome VII from Saccharomyces cerevisiae reveals ARO2, RPL9A, T1P1, MRF1 genes and six new open reading frames.";
 RT Yeast 13:177-182(1997).
 RL Yeast 13:177-182(1997).
 CC -1- FUNCTION: MITOCHONDRIAL PEPTIDE CHAIN RELEASE FACTOR THAT DIRECTS THE TERMINATION OF TRANSLATION IN RESPONSE TO THE PEPTIDE CHAIN TERMINATION CODONS UAA AND UAG.
 CC -1- SUBCELLULAR LOCATION: Mitochondrial.
 CC -1- SIMILARITY: BELONGS TO THE PROKARYOTIC AND MITOCHONDRIAL RELEASE FACTORS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, X60381; CAA42932.1; .

FT REPEAT 2028 2047 2-27.
 FT REPEAT 2051 2070 2-28.
 FT REPEAT 2071 2090 2-29.
 FT REPEAT 2093 2112 2-30.
 FT REPEAT 2120 2139 2-31.
 SQ SEQUENCE 2334 AA; 258329 MW; B75138CCD278BA3 CRC64;

Query Match 5.0%; Score 95; DB 1; Length 2334;

Best Local Similarity 22.1%; Pred. No. 35; Matches 71; Conservative 42; Mismatches 106; Indels 102; Gaps 20;

QY 29 DESITNIFGPRDRNESSPHNIIIN-----HTTA-----YSESHTY-----DMTFELCLIS 74
 DB 260 DSKLDLSEVERSD-KVSKYLEKNEGYLIHLTADENMLKDERYRYPISIDSTLSLS 318
 QY 75 SHNTLNGACPTSENPSSSGEFTNTTLOFTEKRSLIKRELOIKGYKOLLFVSNCPSG 134
 DB 319 SDFVMSAYPTT-NYSASSQKMDAN-----LKAY----- 346
 QY 135 TLNSAFHNCNKNAASGASLYIYPAGELKNLPRGIMDATIKLRVRRYSETYGTITNI 194
 DB 347 VLKGYT-----DKTTGTN-YAFMKFNMLKPIQMVTAKTLKTYVAHSY---YGT----- 393
 QY 195 TIKLTPGKNIQIWLPOFKS---DARYDLNLPRPGGTYIGRNSVDMCFYDGY-----S 244
 DB 394 --KATG-----LMLDTVNSNYDNAKYMTWKPAKSN--IGKADYHKGOMASYDTVAVKS 444
 QY 245 TNSSSLEIFRODNPKSDGKFLYLRKI-----NDDTKELAVYTLISLLAGKSLPT----- 293
 DB 445 WNSGANGYGR-LHTNGCKEYWKKLISANSANKPIEYTYIP-----KGMTPIIKAY 498
 QY 294 -NCTSLNIDAAASLETNMNI 313
 DB 499 HNGDSTGYED-----ISMKV 514

RESULT 13

ID N100_YEAST STANDARD; PRT: 959 AA.

AC 002629;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Nucleoprotein NUP100/NSP100 (Nuclear pore protein NUP100/NSP100).
 GN NUP100 OR NSP100 OR YKL068W OR YKL336.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93054906; PubMed=1385442;
 RA Wente S.R.; Rout M.P.; Blobel G.;
 RT "A new family of yeast nuclear pore complex proteins.";
 RL J. Cell Biol. 119:705-723(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=94378724; PubMed=8091863;
 RA Rasmussen S.W.;
 RT "Sequence of a 20.7 kb region of yeast chromosome XI includes the
 NUP100 gene, an open reading frame (ORF) possibly representing a
 nucleoside diphosphate kinase gene, tRNAs for His, Val and Trp in
 addition to seven ORFs with weak or no significant similarity to
 known proteins.";
 RL yeast 10:569-574(1994).
 CC -!- FUNCTION: ESSENTIAL COMPONENT OF NUCLEAR PORE COMPLEX.
 CC NUCLEOPORINS MAY BE INVOLVED IN BOTH BINDING AND TRANSLLOCATION OF
 CC THE PROTEINS DURING NUCLEOCYTOPLASMIC TRANSPORT.
 CC -!- SUBCELLULAR LOCATION: Nuclear pore complex.
 CC -!- DOMAIN: CONTAINS G-L-F-G REPEATS.
 CC -!- SIMILARITY: BELONGS TO THE GLFG FAMILY OF NUCLEOPORINS.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----

DR EMBL; Z15035; CAA78753.1; -
 DR EMBL; X75780; CAA53406.1; -
 DR EMBL; Z28068; CAA81905.1; -
 DR PIR; B44402; B44402.
 DR PIR; S39173; S39173.
 DR PIR; S44518; S44518.
 DR SGD; S0001551; NUP100.
 DR InterPro; IPR004325; Nucleoporin_Fg.
 DR Pfam; PF03093; Nucleoporin_Fg; 24.
 KW Nuclear protein; Transport; Repeat.
 FT DOMAIN 33 571
 FT SEQUENCE 959 AA; 99988 MW; D3985F9901BBA51 CRC64;

Query Match 5.0%; Score 94.5; DB 1; Length 959;
 Best Local Similarity 19.5%; Pred. No. 12; Matches 80; Conservative 58; Mismatches 157; Indels 115; Gaps 17;

QY 31 SITNIFGPRDRNESS-----PKHNL-----NNHTAYSESHTYDYMTEFLCLSSHTLNG 81
 DB 284 SNGSLFGQNSMNSSTQGVGQNNQNMQINQNNNSNLFGRANTSNASGSLFGQNNQOG 343
 QY 82 ACPTSENPSSSVSG-----ETNITLOFTEKRSLIKRELOIKGY----- 121
 DB 344 SGLEFGQNSQTSQSGSLFGQNNQKOPNTFTQSNIGTL-FQNNNQOQSGSLFGAGPAGT 402
 QY 122 -----QLFKSYNCPGSLTLNSAHFNCKNAASGASLYIYPAGELKNLPRG 168
 DB 403 TGSLEFGQNSSTOPNSLFGTINPTFS-----MTQSQGNSLF--GATKLTMPPG 449
 QY 169 GIMDATIKLRVRRYSETYGTITNITIKLTDKGNIOIWLPOFKS----- 213
 DB 450 GNFTA-----NQSGSNLSLGTKRPASTGSLFGNFTASTYPTPSTNGLFGNNANSTSTTN 505
 QY 214 ---DARYDLNLPRT-GGGTY-----IGRN-----SDVMCTYDGTSTSSSL--- 250
 DB 506 GLFGARPDSQSKPALGGLEFGNSNSNSTIGQMKPVFGTGTONTGLFGATGTNNSAVGSR 565
 QY 251 -EIRPDNNPKSDGKFLYLRKINDTKELAVYTLISLLAGKSLPTNGTSLNIDAAASLETN 309
 DB 566 GKLEFGQNNNTLVNGTQNVPPVNTTQN-----ALGTTAVP-----SLQQAFV-----TN 610
 QY 310 WNRITAVTMP-EISVPLVCPGRIQDAKVENPEAQYMGINIVTTPSS 358
 DB 611 EQLFKISISIPNSITINPVKATTSKVNMADMKRNSLSAYRLAPKPLAPSS 660

RESULT 14

ID LEU2_BUCRP STANDARD; PRT: 471 AA.

AC P48573;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE 3-isopropylmalate dehydratase large subunit (EC 4.2.1.33)
 DE (isopropylmalate isomerase) (Alpha-IPM isomerase) (IPM1).
 GN LEUC.
 OS Buchnera aphidicola (subsp. Rhopalosiphum padi).
 OG Plasmid PRPE.
 OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
 OX NCBI_TaxID=98793;
 RN [1]
 RP SEQUENCE FROM N.A.

MDLINE=95333198; PubMed=7608990;
RA Bricho A.M., Martinez-Torres D., Moya A., Latorre A.;
RT "Discovery and molecular characterization of a plasmid localized in
RT Buchnera sp. bacterial endosymbiont of the aphid Rhopalosiphum
RT padl.";
RL J. Mol. Evol. 41:67-73(1995).
CC -1- CATALYTIC ACTIVITY: 3-ISOPROPYLMALATE = 2-ISOPROPYLMALATE +
CC H(2)O (ALSO CATALYZES 2-ISOPROPYLMALATE + H(2)O = 3-HYDROXY-
CC 4-METHYL-3-CARBOXPENTANONE).
CC -1- PATHWAY: SECOND STEP IN LEUCINE BIOSYNTHESIS.
CC -1- SUBUNIT: CONSISTS OF TWO DIFFERENT SUBUNITS: LEUC AND LEUD.
CC -1- SIMILARITY: BELONGS TO THE ACONITASE/IPM ISOMERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X71612; CAA50617.1; -
DR InterPro: IPR001030; Aconitase.
DR Pfam: PF00330; aconitase; 1.
DR PRINTS: PR00415; ACONITASE.
DR PRODOM: PD000511; Aconitase; 1.
DR PROSITE: PS00450; ACONITASE_1; 1.
DR PROSITE: PS01244; ACONITASE_2; 1.
KW Leucine biosynthesis; Lyase; Iron-sulfur; 4Fe-4S; Plasmid.
FT METAL 347 347
FT METAL 409 409 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT METAL 412 412 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
SQ SEQUENCE 471 AA; 52596 MW; E509ABC23AFCE297 CRC64;

Query Match 5.0%; Score 94; DB 1; Length 471;
Best Local Similarity 22.3%; Pred. No. 5.5;
Matches 71; Conservative 46; Mismatches 98; Indels 104; Gaps 18;

QY 44 SSPKINILNHTAYSESTLYDRMTFLCLSSHTNLNGAC-PTSEPPSSSSVSGENITL 102
DB 101 NHPKGGIV-HVIGPQGTL-PGMTIVCGDSHTHGAFGALSFQIGTSEV-EHVLVT 155
QY 103 QPTEKSLIKRELQIKGYKQ-----LLF-----KSNCPDS 132
DB 156 QTLKQORLKNMKIQTIGKTKNFYAKDITLITIGKGTSGSGGYVEFGVDYIKKNMEE 215
QY 133 GLTNSAHFNCKNNAASGA-----SLYIYPAGELKNLPFGGIMD-----ATLKL 177
DB 216 RMTV-----CNMAIEMGAKSALIAPEDETFYIKGRYS--PGKFWEEAVKYMKTLLT 267
QY 178 RKRRRSE-----TYGT---YTINITIKLPD---KGIOIWLQFQKSDAV 217
DB 268 RSKAFPFDEKFTDVNSLSPQITWGTSPPDVLISINEKIPDFNFKSVKKNLRSACDY-- 325
QY 218 DNLAPPTGGTYIGRNSVDMCFYDGYSTNSSLEIRFQNNPKSDOKFLRKINDTKEI 277
DB 326 -NNKLP---GSLTIDIKIDKFTI-GSCTNSRIEDLSRAKIILKN-----KIKSKIKAI 374
QY 278 AVTSLILLAGKSLTPNGT 296
DB 375 -----LLPQSGT 381

RESULT 15
THBG_SHEEP
ID THBG_SHEEP STANDARD: PRT: 412 AA.
AC P50450;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ThyoXine-binding globulin precursor (T4-binding globulin).
GN SERPINY7 OR TBG.

OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxId=9940;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94192879; PubMed=8143919;
RA "Sheep thyoXine-binding globulin: cDNA sequence and expression.";
RL Mol. Cell. Endocrinol. 98:91-97(1993).
CC -1- FUNCTION: MAJOR THYROID HORMONE TRANSPORT PROTEIN IN SERUM.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: SYNTHESIZED IN THE LIVER; FOUND IN PLASMA.
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR HSSP: P01011; ZACH.
DR InterPro: IPR000215; Serpin.
DR Pfam: PF00079; serpin; 1.
DR SMART: SM00093; SERPIN; 1.
DR PROSITE: PS00284; SERPIN; 1.
KW Serpin; Glycoprotein; Plasma; Signal.
FT SIGNAL 1 16
FT CHAIN 17 412
FT CARBOHYD 35 35
FT CARBOHYD 98 98
FT CARBOHYD 164 164
FT CARBOHYD 252 252
SQ SEQUENCE 412 AA; 45948 MW; E420D3530D1A04D9 CRC64;

Query Match 5.0%; Score 93.5; DB 1; Length 412;
Best Local Similarity 18.4%; Pred. No. 5.1;
Matches 56; Conservative 54; Mismatches 102; Indels 93; Gaps 14;

QY 33 TIIQPPROBNSSPKNINLNIHTAYSES-----HTLYDRMFLCLSSHTNLNG 81
DB 155 TEVFSTDFSNVSAAOEI-NSHVERQTKGIYGLIQLDKPNTITVLYVNLCKKA-QMANP 212
QY 82 ACPTSENPSSSVSGENTITLOFTEKSLIKRELQIKGYKQLPFKSVNCPGSLTNSAHF 141
DB 213 FDPKSTEESSFLVDKTT-TVQ-----VPMHQVDQYIHLVDIELNC-----TVLQM 258
QY 142 NCKNNAASGASLYIYPAGELKNLPFGGIMDATALRKVRYSERYGYTINITIKLPDK 201
DB 259 DYSKNA---LALFVLPKEGMEGV-----EGAMSSKILKKNV-----RLQK 297
QY 202 GNIQIWLQFQKSDAVDNLRFPGG-----GTIYGRNSVDM 237
DB 298 GWNVLFVPRFISATYD-----GGILLKMGIDAPADNADFSGLTKDNGLKVSNAVAKA 352
QY 238 CFYDGYSTNSSLEIRFQNNPKS-----DGKFLYRKINDTKEIAYTSLILLAGK 288
DB 353 MYIIGKGTGALPEVRF-LNQPETILLHPIIOFDSFLLILKMKR-----SLTFQK 405
QY 289 SLTPT 293
DB 406 VVDPT 410

Search completed: July 1, 2002, 11:25:52
Job time: 128 sec

This Page Blank (uspto)

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 11:23:04 ; Search time 28.51 Seconds
(without alignments)
2190.501 Million cell updates/sec

Title: US-09-839-894-10
Perfect score: 1886
Sequence: 1 MNKILFFTLFFSSVLETFEA.....EAGQYMGNTVTFPPSSQTL 361

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SPTREMBL_19:*

- 1: sp.archaea:*
- 2: sp.bacteria:*
- 3: sp.fungi:*
- 4: sp.human:*
- 5: sp.invertebrate:*
- 6: sp.mammal:*
- 7: sp.mhc:*
- 8: sp.organelle:*
- 9: sp.phage:*
- 10: sp.plant:*
- 11: sp.rodent:*
- 12: sp.virus:*
- 13: sp.vertibrate:*
- 14: sp.unclassified:*
- 15: sp.virus:*
- 16: sp.bacteriap:*
- 17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1886	100.0	361	2	093G67
2	949.5	50.3	363	2	047115
3	930.5	49.3	364	2	047119
4	275.5	14.6	359	2	09XDS1
5	115.5	6.1	847	2	P71132
6	110	5.8	846	2	P71133
7	109	5.8	2349	2	P94750
8	109	5.8	2383	16	P76347
9	108	5.7	872	2	09RLA0
10	108	5.7	2529	16	025579
11	105.5	5.6	430	3	096VN2
12	104.5	5.5	1441	16	09CFL1
13	104	5.5	803	10	09AUN5
14	103.5	5.5	691	16	09RZS7
15	103.5	5.5	1269	17	097Z06
16	103.5	5.5	2399	16	09ZKS9

17	102.5	5.4	1736	5	095PH7
18	102.5	5.4	2747	2	091800
19	102	5.4	692	17	097YM6
20	100	5.3	1014	2	046962
21	100	5.3	1752	10	09LX60
22	99.5	5.3	1300	16	09CK27
23	99.5	5.3	5627	16	091120
24	99	5.2	456	2	048508
25	99	5.2	802	16	092C16
26	99	5.2	839	2	P7792
27	99	5.2	856	16	09PHX4
28	98.5	5.2	4919	2	09ZHL0
29	98	5.2	587	16	092E42
30	98	5.2	1014	2	047106
31	97.5	5.2	855	15	066861
32	97.5	5.2	1755	17	026812
33	97	5.1	334	2	093QW7
34	97	5.1	454	10	081310
35	97	5.1	803	2	093430
36	96.5	5.1	2364	2	046342
37	96	5.1	406	2	09FOR4
38	96	5.1	430	10	093ZNS
39	96	5.1	992	8	021375
40	95.5	5.1	895	10	080973
41	95.5	5.1	2081	17	09HJ37
42	95	5.0	198	2	09K541
43	95	5.0	561	16	09K6M4
44	95	5.0	867	2	093G68
45	94.5	5.0	200	10	09SB15

ALIGNMENTS

RESULT 1
ID 093G67 PRELIMINARY; PRF; 361 AA.
AC 093G67;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE MINOR PILIN PROTEIN CSAE.
GN CSAE.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=El1881A;
RA Althoum Z.D., Levine M.M., Galen J.E., Barry E.M.;
RT "Isolation and characterization of EPEC CSA fimbriae encoding genes,
RT and their expression in Shigella flexneri 2a guinea strain CVD 1204.";
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF296132; AAK97137.1;
SQ SEQUENCE 361 AA; 40102 MW; 2E1E74ABDB0EB353 CRC64;

Query Match	Best Local Similarity	Score	1886;	DB 2;	Length	361;
Matches 361; Conservative	0;	Mismatches	0;	Indels	0;	Gaps 0;
OY 1 MNKILFFTLFFSSVLETFEAVSADKIPGDEISITNFGPRDRNESSPKHNILNNHTAYSE 60						
DB 1 MNKILFFTLFFSSVLETFEAVSADKIPGDEISITNFGPRDRNESSPKHNILNNHTAYSE 60						
OY 61 SHLYDRATFLCSSAHNTLACPTSENSSSVSGETNITLQFTKRSIKRELQIKGY 120						
DB 61 SHLYDRATFLCSSAHNTLACPTSENSSSVSGETNITLQFTKRSIKRELQIKGY 120						
OY 121 KOLFESVNCPSGLTNSAHFNCKNNAASGASLYLTPAGELKNLPGGIMDADLKLRYK 180						
DB 121 KOLFESVNCPSGLTNSAHFNCKNNAASGASLYLTPAGELKNLPGGIMDADLKLRYK 180						

```

QY 181 RRYSEYGVYTIINTIKLTDKGNIOIWLPOFKSDARVDLNRPTGGTYIGRNSVDMCFY 240
    |||||:|||||:|
Db 181 RRYSEYGVYTIINTIKLTDKGNIOIWLPOFKSDARVDLNRPTGGTYIGRNSVDMCFY 240
QY 241 DGYSTNSSLEIRFODNNPKSDGKFYLRKINDOTKEIAYTLLSLACKSLPTNGTSLNT 300
    |||||:|||||:|
Db 241 DGYSTNSSLEIRFODNNPKSDGKFYLRKINDOTKEIAYTLLSLACKSLPTNGTSLNT 300
QY 301 ADAASLETNNRITAYTAVMEISVAVLCWPGRLQDAKVENPEAGQYMGINIVTFPTSSQT 360
    |||||:|||||:|
Db 301 ADAASLETNNRITAYTAVMEISVAVLCWPGRLQDAKVENPEAGQYMGINIVTFPTSSQT 360
QY 361 L 361
Db 361 L 361

RESULT 2
ID 047115 PRELIMINARY; PRT; 363 AA.
AC 047115;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
DE COOD PRECOURSOR.
GN COOD.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94344028; PubMed=7915003;
RA Froehlich B., Karakashian A., Melsen L.R., Wakefield J.C., Scott J.R.;
RT "Cooc and Cood are required for assembly of CSI pill."
RL Mol. Microbiol. 12:387-401(1994).
DR EMBL: X76908; CAA54230.1;
DR InterPro: IPR001412; trna-synt.L.
DR PROSITE: PS00178; AA_TRNA_LIGASE_I; UNKNOWN_1.
KW Signal.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 363 COOD.
SQ SEQUENCE 363 AA; 40139 MW; AA9488AADD5F72 CRC64;

```

Query Match 50.3%; Score 949.5; DB 2; Length 363;
 Best Local Similarity 53.1%; Pred. No. 6.4e-66;
 Matches 197; Conservative 55; Mismatches 98; Indels 21; Gaps 11;

```

QY 3 KIIFLFTLFSSVLTFAVSADKIPGDE--STINIF-GPR-DRRESSPKNINLNNHITAY 58
    |||||:|||||:|
Db 2 KIIFLFTLFSSVLTFAVSADKIPGDE--STINIF-GPR-DRRESSPKNINLNNHITAY 57
QY 59 SESHTLVDRMTFLCLSSHNTLNGACPTSENPSSSVS--GETNITLQTEKRSILKRELQI 117
    |||||:|||||:|
Db 58 SLHSLYDRIVFLTSSSNVNGACPTI---GTSQVOYGTITITLOTEKRSILKRNINL 114
QY 118 KGYKOLLFKSVNCPG--LTLSNAHFNCNNA--ASGASLYLYIPAGELKMLPFGIWDAT 174
    |||||:|||||:|
Db 115 AGKKKPIWENOSCDFSNMLVNSKSGAHNAGTLLNLYIPAGEINKLPFGIWEAT 174
QY 175 LKLRVRRYSET---YGTITNITIKLTDKGNIOIWLPOFKSDARVDLNRPTGGTYI 230
    |||||:|||||:|
Db 175 LKLRVRRYSET---YGTITNITIKLTDKGNIOIWLPOFKSDARVDLNRPTGGTYI 230
QY 231 GRNSVDMCFYDGYSTNSSLEIRFODNNPKSDGKFYLRKINDOTKEIAYTLLSLACKSL 290
    |||||:|||||:|
Db 234 GSNSLDMCFYDGYSTNSSLEIRFODNNPKSDGKFYLRKINDOTKEIAYTLLSLACKSL 292
QY 291 TPTNGSTSLNADASLETNNRITAYTAVMEISVAVLCWPGRLQDAKVENPEAGQYMGIN 350
    |||||:|||||:|
Db 293 YPVGSGFTINDSSVLETNNRITAYTAVMEISVAVLCWPGRLQDAKVENPEAGQYMGIN 352

```

```

QY 351 NYTFPTSSQTL 361
    |||||:|||||:|
Db 353 YTFPTSSQTL 363

```

```

RESULT 3
ID 047119 PRELIMINARY; PRT; 364 AA.
AC 047119;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
DT 01-NOV-1998 (Tremblrel. 08, last annotation update)
DE COOD PRECOURSOR.
GN COOD.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C91F-6;
RX MEDLINE=96071908; PubMed=7591145;
RA Froehlich B.J., Karakashian A., Sakellaris H., Scott J.R.;
RT "Genes for CS2 pill of enterotoxigenic Escherichia coli and their
interchangeability with those for CSI pill."
RL Infect. Immun. 63:4849-4856(1995).
DR EMBL: Z47800; CAA87763.1;
KW Signal.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 364 COOD.
SQ SEQUENCE 364 AA; 40979 MW; EC7AC738A0E195B6 CRC64;

```

Query Match 49.3%; Score 930.5; DB 2; Length 364;
 Best Local Similarity 48.4%; Pred. No. 1.9e-64;
 Matches 177; Conservative 64; Mismatches 118; Indels 7; Gaps 3;

```

QY 1 MNKILFTLFSSVLTFAVSADKIPGDESTINIFGPRDRN--ESSPKNINLNNHITAY 58
    |||||:|||||:|
Db 1 MNKILFTLFSSVLTFAVSADKIPGDESTINIFGPRDRN--ESSPKNINLNNHITAY 57
QY 59 SESHTLVDRMTFLCLSSHNTLNGACPTSENPSSSVS--GETNITLQTEKRSILKRELQI 118
    |||||:|||||:|
Db 61 NKDHSLEDRMTFLCLSSHNTLNGACPTSENPSSSVS--GETNITLQTEKRSILKRELQI 118
QY 119 GYKOLLFKSVNCPG--LTLSNAHFNCNNA--ASGASLYLYIPAGELKMLPFGIWDAT 175
    |||||:|||||:|
Db 119 GYKOLLFKSVNCPG--LTLSNAHFNCNNA--ASGASLYLYIPAGELKMLPFGIWDAT 178
QY 176 KLRVRRYSETGYTIINTIKLTDKGNIOIWLPOFKSDARVDLNRPTGGTYIGRNSV 225
    |||||:|||||:|
Db 179 ELKRVRRYSETGYTIINTIKLTDKGNIOIWLPOFKSDARVDLNRPTGGTYIGRNSV 228
QY 236 DMCFYDGYSTNSSLEIRFODNNPKSDGKFYLRKINDOTKEIAYTLLSLACKSLPTNG 295
    |||||:|||||:|
Db 239 EMLYDGYSTNSSLEIRFODNNPKSDGKFYLRKINDOTKEIAYTLLSLACKSLPTNG 298
QY 296 TSLNADASLETNNRITAYTAVMEISVAVLCWPGRLQDAKVENPEAGQYMGINIVTF 355
    |||||:|||||:|
Db 299 EAFITNDSSLEIFNMNRKISVLSIPISIPVLCWPGRLQDAKVENPEAGQYMGINIVTF 358
QY 356 PSSQTL 361
Db 359 PSSSSL 364

RESULT 4
ID 09XDS1 PRELIMINARY; PRT; 359 AA.
AC 09XDS1;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, last sequence update)

```


RC SEQUENCE OF 2118-2383 FROM N.A.
RA STRAIN-K12:
RA Alba H., Baba T., Fujita K., Hayashi K., Honjo A., Horuchi T.,
RA Ikemoto K., Inada T., Isono K., Itoh T., Kanai K., Kasai H.,
RA Kasimoto K., Kim S., Kimura S., Kitagawa M., Kitakawa K., Makino K.,
RA Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y.,
RA Nishimoto H., Nishio Y., Oshima T., Saito N., Sempel G., Seki Y.,
RA Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano M.,
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AE000289; AAC75042.1; -
DR EMBL: D90837; BAA15800.1; -
DR InterPro: IPR003344; Big_1.
DR InterPro: IPR003335; Intimin.
DR InterPro: IPR002482; LysM.
DR InterPro: IPR00601; PKD_domain.
DR Pfam: PF02369; Big_1; 13.
DR PRINTS: PR01369; INTIMIN.
DR SMART: SM00257; LysM; 1.
DR SMART: SM00089; PKD; 4.
KW Complete proteome.
SQ SEQUENCE 2383 AA; 251391 MW; 129975F4CC64225C CRC64;

Query Match 5.8%; Score 109; DB 16; Length 2383;
Best Local Similarity 20.0%; Pred. No. 14;
Matches 78; Conservative 65; Mismatches 157; Indels 90; Gaps 15;

QY 13 SSYLFTEFAVSADKIPGDESITNIFGRDRNESSPKHNLNNHTAY---SESHLYDRMT 69
DB 762 SAKIATILASNNQVLNEMANANTSVNVADEGS---NPINDHTVTFVAVLGSGATSPNNQ 818
QY 70 FLCLSSHTNLNACPTSENPSSSVSGETNITLQTFEKSLSL-----K 112
DB 819 ---TAKTGVNGLA--TFDLKSSKQEDNTVEVLENGVKQTLVSEFGDSSTAQVDLQSK 873
QY 113 RELQIRGYKQL-LFKSVNCPGSLTNSAHFNCKNNAASGSLVLYIPAGELKLNLPFGI 171
DB 874 NEVVADGNSVMTATVRBAKGNLMDVAVTFVNVSAEAKLSLSEVNSHD-----GIA 926
QY 172 DATLKLVRKRRSEYGYTITITIKLTDKGNIQIWLPOFKSDARVDLMLRPTGGGTG 231
DB 927 TATLT-----SLKNGDYRVTFASVSSGSAQNVNFIGDQSTALTLV--PGDITV-- 976
QY 232 RNSVDMCFDGIYSTNS---SLEIFQDNPKSDCKFLIRKINDTKELATVLSLLAG 287
DB 977 -----TNTAPQYMTATLQDKNGNPLKDKETFEFVSPND---VASKFSISNG 1019
QY 288 KSLPTNGTSL-----NTADAASLETNNRITVATVMPESVPLC 327
DB 1020 KGMTDNGVAIASLITGLAGTHIMAKRLANSNVSDAQPMTFVADKDRAYVVLQTSKAEII 1079
QY 328 WPG--RLQDPAKYENPEAQYMGMINVTF 355
DB 1080 GNGVDFTTLATVYKDP-SNHVPAGITVNF 1108

RESULT 9
Q9RLA0 PRELIMINARY; PRT; 872 AA.
AC Q9RLA0:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE DNA POLYMERASE I.
GN POLA.
OS Rickettsia typhi.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiiae; Rickettsia.
OX NCBI_Taxid=785;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WILMINGTON;
RX MEDLINE=99416441; PubMed=10486973;

RA Andersson J.O., Andersson S.G.E.;
RT "Genome degradation is an ongoing process in Rickettsia.";
RL Mol. Biol. Evol. 16:1178-1191(1999).
DR EMBL: AJ238757; CAB56108.1; -
DR HSSP: P00582; 1KFD.
DR InterPro: IPR002421; 5.3_exonuclease.
DR InterPro: IPR002298; DNA_pol_A.
DR InterPro: IPR001098; DNA_pol_A.
DR InterPro: IPR000513; Exo_N1.
DR InterPro: IPR003584; HH_2.
DR Pfam: PF01367; 5.3_exonuclease; 1.
DR Pfam: PF02739; 5.3_exonuc N: 1.
DR Pfam: PF00476; DNA_pol_A: 1.
DR PRINTS: PR00868; DNAPOL1.
DR SMART: SM00475; 53EXOC; 1.
DR SMART: SM00279; HNH2; 1.
DR SMART: SM00482; POLAC; 1.
DR PROSITE: PS00447; DNA_POLYMERASE_A; 1.
SQ SEQUENCE 872 AA; 99309 MW; BC2BEE3490879A08 CRC64;

Query Match 5.7%; Score 108; DB 2; Length 872;
Best Local Similarity 20.4%; Pred. No. 4.6;
Matches 77; Conservative 50; Mismatches 139; Indels 112; Gaps 15;

QY 2 NKILFFTEFSSVLTFAVSADKIPGDESI-----TNIQPRDRNESSPKH 48
DB 383 NKSIRKITYSLKHLFEVANSKIRAIEDLELMQYTLNAGLVQKNLFA-----ETLTKD 437
QY 49 NILNNHTVASESHLYDRMTFLCLSSHNTLN-----GACPTSENPSSSVSGETNIT 101
DB 438 NITNESAKYINFTSLY-KOTLLALQKKAFRLYREIDLPCTIILDKMEKIGIKVANYI 496
QY 102 LQFTEK--NSLIRELQI-----KCYKQLFKSVNCPGSLTNSAHFNCKNN 146
DB 497 HQLSDFGEFELIKIEEIFALSGTKFNINASQKLSLELFKMKQLPSGNFLAKTSYSTK- 555
QY 147 AASGASLYLTPAGELKNLPFGIWDATLKL-----VKRRSEYGYTITITIKL-- 198
DB 556 -----AGILKISEDDYHATLTLRLRQTLKNTYDLSLPROINNITRIRHT 603
QY 199 -----TDKGNIQIWLPOFKSDARVDLNLPTGGGTGYIGNSVDMCFY--DGY---STNS 247
DB 604 TFLQTTTTRGLSSQEPNIO-----NIPTRSSG-----NKIRQAFIADGKILSADY 652
QY 248 SLEIRF-----QDNPKSDCKFLIRKINDTK-----IAY 279
DB 653 SQELRLILSHIANVDYLKQAFINKEDIHQTAQIFNLQKHELTSERRKAKAINFGIIT 712
QY 280 TISLLAGKSLPTNGTS 297
DB 713 GISAFGLAQNLVNSGTA 730

RESULT 10
O25579 PRELIMINARY; PRT; 2529 AA.
AC O25579;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE TOXIN-LIKE OUTER MEMBRANE PROTEIN.
GN HP0922.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_Taxid=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kertlavage A.R., Clayton R.A., Sutton G.G.,
Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,

RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Lotius B., Richardson D., Dodson R., Khalak H.G., Glodok A.,
RA McEnany K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Goayave J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weiman J.M., Fujii C., Bowman C., Matthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen *Helicobacter*
RT *pylori*.";
RL Nature 368:539-547(1997).
DR EMBL: AE000602; AAD07969.1; -
DR TIGR: HP0922; -
DR InterPro: IPR004311; VACA.
DR Pfam: PF03077; VACA.3.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 2529 AA; 274561 MW; 440882E8644472EC CRC64;

Query Match 5.7%; Score 108; DB 16; Length 2529;
Best Local Similarity 22.5%; Pred. No. 17;
Matches 84; Conservative 50; Mismatches 122; Indels 118; Gaps 22;

OY 85 TSENSSSSVSEETITLOFTEKRSLL-----KRELQIKYKOLLER---SVNCP 131
DB 977 TSNFATYQLLGNTFTL---SSQSILNFGDTTLQNNANITLGNKSAFNSLTLDMN 1033
OY 132 SGLTL-NSAHFNCNKAA--SGASLYLYPA-GELKNLPFGGIMDATLKLVRKRYSEY 187
DB 1034 SNLSLNDQSVLANNTSAFNNQASLNIYNGSOATPNSLFPNG---GTLSLNASSKINASN 1090
OY 188 GYTYITITIKLND-----KGNIQIWLPOKSDARVDLNLPTGGGTYYIGRNSVD 236
DB 1091 ASFSNNTTINLDDSVLASNTSLNANI-----NFGASQADF---CGNTIIDTASFN 1140
OY 237 MCFYGGYSTNSSL---EIRFDNNPK-----SDGKFLRKIN---DDT 274
DB 1141 --FDSASSLNFNNLTANGALNNGYTPSLTKALMSVSGQFVLGNMGDILMSDINFENIT 1198
OY 275 KEIATYTLILLAGKSLTPTNGTS-----LNTADAASLEFN---WNRITAVTMPET 321
DB 1199 KSVY--NLTNOKGKITGSGANGYEKILFYGMKTIQNAISNNNIQTWFSFNLPLNSSOI 1256
OY 322 -----SVPVL-----CWPGRLODAKVENP-----EAGGY--M 347
DB 1257 IQESIKNGDLTIEVLNPNNSASNTIFNIAPELYNQASQKQNTGVSYDSDMQAGTYVLT 1316
OY 348 GNINVTFTF-SSQF 360
DB 1317 SNIKGLFTPKGSQT 1330

RESULT 11

O96VN2 PRELIMINARY: PRT; 430 AA.
AC O96VN2;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DI 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
GN EXTRACELLULAR ALKALINE PROTEASE.
OS Blumeria graminis.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomycetes;
OC Erysiphales; Erysiphaceae; Blumeria.
OX NCBI_TaxID=34373;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhang Z., Gurr S.J.;
RT "An extracellular alkaline protease from *Blumeria graminis*.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY039006; AAK84436.1; -
KW protease.
SQ SEQUENCE 430 AA; 44817 MW; 89B4D822BBBC3CB5 CRC64;

Query Match 5.6%; Score 105.5; DB 3; Length 430;
Best Local Similarity 21.7%; Pred. No. 3;
Matches 86; Conservative 48; Mismatches 128; Indels 135; Gaps 21;

OY 7 ITLFFSSVLEFFAVSADKI---PDSEITNIFSPDRNESPKNH--ILNNHTIAYS-- 59
DB 12 VITLY---LFSFAYSAPPLFSPANPPSIYGLDLPISLNDAPAEIMANRYIVYVSKD 67
OY 60 -----ESHITLDRMTPLCLSSHNTLN-----GACPTSENSS 91
DB 68 ATDDAVESH---QSTIKTALKKSLNATSIDGRLSPMHTFKKGMKGLDDEDMT 123
OY 92 SSVSEETITLOFTEKRSLLIKRELQIKYKOLLERFSVNCPSGLT-LSAHFNCNK----- 145
DB 124 IIEEASEV-----SVLEADTQV-GLELT-EQIEAPIGLVLSHAKSESEYFED 173
OY 146 NASGASLYLY-IPAGELKNLPFGGIMDATLKLVRKRYSEYTYTINITIKLTDKNT 204
DB 174 NASDAGIYGYIDGIRASHFEG-----RATMKAN- 205
OY 205 QTWLQFQKSDARVDLNLKPTGGGTYYIGRNSVDMCFYDGTNSSLLEIRFDNNPKSDGK 264
DB 206 -----FIDIDINEDON---GHSVATTIGASF--GVAKNIELVGIKVLDAKKGNSA 253
OY 265 FYLRKINDTKEIATYTLILLAGKSLTPTNGSLNIADAASLETNMNRITAVTMPETISVP 324
DB 254 NVLRGVNFDVNERG-----LAGKAV-----VNISIGSKSKALN-TAIEALTKAGTYV 301
OY 325 VICWPGRLQDAKVENPEAGQYMNIN--VTFPSS 358
DB 302 VAA-----GNSNKDATSFSPAS 318

RESULT 12

O9CFL1 PRELIMINARY: PRT; 1441 AA.
AC O9CFL1;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DI 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
GN UNKNOWN PROTEIN.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=1137471;
RA Bolotin A., Wincker P., Manger S., Jallion O., Malarre K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium *Lactococcus*
RT *lactis* ssp. *lactis* IL1403.";
RL Genome Res. 11:731-753(2001).
DR EMBL: AE006376; AAK05556.1; -
DR InterPro: IPR001899; Gram_pos_anchor.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 1441 AA; 142946 MW; 4CD4ABC561BA15EA CRC64;

Query Match 5.5%; Score 104.5; DB 16; Length 1441;
Best Local Similarity 21.4%; Pred. No. 16;
Matches 91; Conservative 59; Mismatches 160; Indels 115; Gaps 21;

OY 10 IFFSVSLFTF-----AVSADKI-PG-----DESTINIFGPRDRNESPKNH----- 50
DB 28 LYASSVLLIYVGAIAAPSVKADQVLPQGGQGTGANTATGALGASNAAGSAGSNNVSLAN 87
OY 51 -LNNHTIYVSESHLYLDRMTPLCLSSHNT--LNGACPTSENDS----- 91

```

Db 88 SANN--VNYSSAOSAAGSVYSSAASSAVASSNPLSANDPSYTDGSVNNSNQLKT 145
QY 92 -----SSVSGENTIT-----LQTEKRLIKRELQIKGKOLLFKSVNPSGLT 135
Db 146 DSGVENSATNTNSYIVODLNSSGVAGTATGSSAQSASLATPGSSDALKT-----SDTN 201
QY 136 LNSAHFNCKNNAASGASLYLYIPAGELKNLPGGIMDATLKLVRKRYSEFGTYTINIT 195
Db 202 LNS---NCAVYAGSAGSSVSFDIKDIOQVPGA-----LNIYNGATGSI-TWTVSNP 249
QY 196 IKLTKG-----NIQIMLPQFKSDARVDLN---LRPTGGCTY-IGRNSVDMCFYDGYSTN 246
Db 250 SDFTNAGLSVANVAVYDASKGDCALNVNGNEAYQTAGGVYIGTENY-----PGYPTG 304
QY 247 SSSLEIRFQDNKPK-----SDGKFYLRKINDPTKEIAYTSLSLAGSLRPT 293
Db 305 SSASMEPTTESASIMASASSVAASYSDANITALASOYADASTL-YNSYAAANGDGTAA 363
QY 294 NQTSINLIDAAASLETNMRITAVTPEISVPLVPCGRLOLDAKVENPEAGGYMGINVT 353
Db 364 SYANNLVAIQASIASNNNAIALDIGST-----PGYLR-----SSVASQAV-VNIT 408
QY 354 FTPSS 358
Db 409 YTTDS 413

```

RESULT 13

```

ID 09A0U5 PRELIMINARY; PRT: 803 AA.
AC 09A0U5:
DT 01-JUN-2001 (TREMBLREL. 17, Created)
DT 01-JUN-2001 (TREMBLREL. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLREL. 19, Last annotation update)
DE HYPOTHEETICAL 89.2 KDA PROTEIN.
GN OSJNBA0040E01.17.
OS Oryza sativa (Rice).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
OC Spermatophyta: Magnoliophyta: Liliopsida: Poales; Poaceae;
OC Eriatolidae: Oryzae; Oryza.
OX NCBI_Taxid=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. NIPONBARE;
RA Busell C.R., Yuan Q., Ouyang S., Moffat K.S., Hill J.N., Burr P.C.,
RA Hsiao J., Zismann V., Pal G., Bowman C.L., Fujii C.Y., Vanaaken S.E.,
RA Bowman C.L., Craven B., Utterback T.R., Khalak H., Feldblyum T.V.,
RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
RA "Oryza sativa chromosome 3 BAC OSJNBA0040E01 genomic sequence.",
RL Submitted (AUG-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AC079887; AAK16187.1; -.
KW Hypothetical protein.
SQ SEQUENCE 803 AA; 89209 MW; E119841096A8BDAC CRC64;

```

```

Query Match 5.5%; Score 104; DB 10; Length 803;
Best Local Similarity 20.2%; Pred. NO. 8.5; Indels 68; Gaps 10;
Matches 65; Conservative 44; Mismatches 145;
QY 33 TNIF-GPRDRNESSPKH-----NILNNH-ITAVSESHLYDRMTFLCLSSH 76
Db 406 TNIIDGOSLAKTYISHGCVATNGVMKGRDVMENHPLGYSSSHAHHPQWELPMTCST 465
QY 77 NTLGACCTSENPSSSVSGETNITLQTEKRLIKRELQIKGYOLLFKSVNCPDGLTL 136
Db 466 HLLPPLPMISNNPNSSFRGTGLKAEMLGAYSTENTENLIDNSER---KGLICPKSSEQ 522
QY 137 NSAHFNCKNNAASGASLYLYIPAGEL-----KNLPRGCT---WDATIKRVRKRYSE 185
Db 523 NVAENICKAAEYOCNDISQIVCNOOYVLLPLNKASHFGGLPTKFKDGLVSRQKRRRA 582
QY 186 TYGVTYITITIKLDKGNIQIMLPQFKSD-ARVDLNLRTPTGGCTYIGNSVDMCFYDGY 244

```

```

Db 583 T-----SLLSWHAQVMGSGSKTHHTRKPELDMAHATRLRYEKVDAENMK 626
QY 245 TSSSLEIRFQDNKPKSDGKFYLRKINDPTKEIATLSSLAGKSLPTNGSLINADAA 304
Db 627 TKNSTF-----VSOAQKRLAFTTKIMQYILPVLPRLLAANAID-- 665
QY 305 SLETNMRITAVTPEISVPL 326
Db 666 SCETIYVIRSRALPDARNPAT 687

```

RESULT 14

```

ID 09RZS7 PRELIMINARY; PRT: 691 AA.
AC 09RZS7:
DT 01-MAY-2000 (TREMBLREL. 13, Created)
DT 01-MAY-2000 (TREMBLREL. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLREL. 17, Last annotation update)
DE HYPOTHEETICAL 69.7 KDA PROTEIN.
GN DRB0037.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_Taxid=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-R1.
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1."
RL Science 286:1571-1577(1999).
DR EMBL; AE001826; AAF12628.1; -.
DR TIGR; DRB0037; -.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR001899; Gram_pos_anchor.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Hypothetical protein; Plasmid; Complete proteome.
SQ SEQUENCE 691 AA; 69702 MW; 744112A6a23C022E CRC64;

```

```

Query Match 5.5%; Score 103.5; DB 16; Length 691;
Best Local Similarity 21.3%; Pred. NO. 7.7; Indels 97; Gaps 15;
Matches 70; Conservative 40; Mismatches 121;
QY 28 GDESTINIFGPRDRNESSPKHNLNNHITAVSESHLYDRMTFLCLSSHNTLNGACPTSE 87
Db 218 GSPFLTNVVPSSPNPLPPTG-----TTPGCTGTGVGSGN 253
QY 88 NESSSVSGEINI-TLQTEKRLIKRELQIKGYOLLFKSVNCPDGLTNSAHFNCKN 146
Db 254 GFAESTGPAVYTGISTGSTOPVVDASGNOYAYPKADAEFVN-PDVTMTIITNPPT 312
QY 147 AASGASLYLYIPAGELKNLPGGIMDATLKLVRKRYSEFGTYTINITIKLTD-KGNQ 205
Db 313 AAT-----YELV--DLGSLPAG-----YVFTTDANGNP- 340
QY 206 IWLPOFKSDARVDLNLRTPGG-GTYIGRNSVDMCFYDGYSTNSSLEIRFO---DNPKS 261
Db 341 -LPPTDNGREVTYAGNGGTATY-----RVVYVTPDRESAAAVAPRIRIPGVGDN--- 391
QY 262 DGKFLYLRKINDPTKEIATLSSLAGKSLPTNGSLINADAAASLETNMRITAVTPEI 321
Db 392 -----RGIYDATYTYNVLISNLFKGNFTALGVSDVPTVRT-----VTOPSV 435
QY 322 SVPLVPCWPGRLQDAKVENPEAGGYMGN 349

```



```

51 TACATTTGCTGTATCGCAGATAAATTCGCCGAGATGAAGCATATCA 100
   :      :      :      :      :      :      :      :
17 rGlyInserThrPheIsthrAsnValGlnAlaGlySerIleAsnLysThr 34
101 ATATTTTGGCCCGCGTGACGAGAAC.....GAATCTTCCCCCAACAT 144
   :      :      :      :      :      :      :      :
34 InsetIleGlyProIleAspArgSerAlaAlaIleSerTyProAlaHis 50
145 AATATTTAAATACCATATTTACAGATACAGGAAAGTCATCTGTGA 134
   :      :      :      :      :      :      :      :
51 TyIlePheHisGlnHisValAlaGlyTyAsnLysAspHisSerLeuPhe 67
195 TGATAGGATGACTTTTATGTTGTCTTCTCACAATACACTTAATGAG 244
   :      :      :      :      :      :      :      :
67 eaSpArgMetThrPheLeuLysMetSerSerThrAspAlaSerLysGly 84
245 CATGCCAACCCAGTGAAGATCTCAGCATTCATCGCTCAGCGGTGAACA 294
   :      :      :      :      :      :      :      :
84 lacYProThrGlyGlnAsnSerLysSerSer.....GlnGlyGluThr 98
295 AATATACATTTACATTTACGGAAGAAAGATTAATAAAGACAGCT 344
   :      :      :      :      :      :      :      :
99 AsnIleLysLeuIlePheThrGlnLysLysSerLeuAlaArgLysThrIe 115
345 ACAAAATTAAGGCTATTAACAAATTTATGTTCAAAAGTGAAGTTCG... 390
   :      :      :      :      :      :      :      :
115 uasLeuLysGlyTyLysArgPheLeuTyrgLuserAspArgCysIleH 132
391 ....CCATCCGGCCTAACACTTTAACTCAGCTCATTTTAACTGTAATAA 435
   :      :      :      :      :      :      :      :
132 IstTyValAspLysMetAsnLeuAsnSerHisThrValLysCysValGly 148
436 AAGCGGCTCAGTGAAGTTATATTTATATTCCTCCGCGGCAACT 485
   :      :      :      :      :      :      :      :
149 SerPheThrArgGlyValAspPheThrLeuTyIleProGlnGlyLys 165
486 AAAAAATTTGCTTTGTGTGTATCTGGATCTGAGTCTGAACTGAAGTAAAG 535
   :      :      :      :      :      :      :      :
165 eaSpGlyLeuLeuThrGlyGlyIleThrProGlnAlaThrLeuGlnLeuArg 182
536 TAAAAACACCATATAGTACGACCTATGGAACCTTACACATTAATATACAT 585
   :      :      :      :      :      :      :      :
182 allysArgHisTyAspTyAsnHisGlyThrTyLysValAsnIleThr 198
586 ATTAATTAAGTATGATTAAGGAAATATTCAGATATGTTACCTAGTTCACA 635
   :      :      :      :      :      :      :      :
199 ValAspLeuThrAspLysGlyAsnIleGlnValIlePrThrProLysPheH 215
636 AAGTGAAGCTCGCGTGCATCTTAACCTTGCCTCCACTGTGGGGGACAT 685
   :      :      :      :      :      :      :      :
215 sSerAspProArgIleAspLeuAsnLeuArgProGlnGlyAsnGlyLysT 232
686 ATATTGGAAGAAATTTGTTGATATGCTTTTATGATGATATAGTACT 735
   :      :      :      :      :      :      :      :
232 ySerGlySerAsnValLeuGlnMetCysLeuTyAspGlyTySerThr 248
736 AACAGCAGCTCTTTGGAGATTAAGATTTCCAGATTAACAATCCCTAAATCTGA 785
   :      :      :      :      :      :      :      :
249 HisSerGlnSerIleGlnMetArgPheGlnAspAspSerGlnThrLys 265
786 TGGGAAATTTTATCTAAGGAAATAATATGATGACCAACAAAGAAATTCAT 835
   :      :      :      :      :      :      :      :
265 nasGlnTyAsnLeuIleLysThrGlyGluProLeuLysLysLeuProt 282
836 ATACTTTGACCTTCTTGGCGGGTAAAGTTTAACTCCAAACAATGGA 885
   :      :      :      :      :      :      :      :
282 yLysLeuSerLeuLeuGlyGlyArgGlnPheTyProAsnAsnGly 298
886 AGGTATTAATATGCTGACGAGCTCTCTGGAACAACAACAGCAATAG 935
299 GlnAlaPheThrIleAsnAspThrSerSerLeuPheIleAsnTrpAsnTr 315
936 AATTACAGCTGTACCATGCCAGAAATCAGTTCGCGTGTGTGTGGC 985

```

```

||||| .....|.....|.....|.....|.....|.....|.....|.....|
315 gIleLysSerValSerLeuProGlnIleSerIleProValLeuCysTrp 332
986 CTGGACGTTTGCATTTGATGACAAAGTGAAATCCGAGCTGACAA 1035
   ||.....|.....|.....|.....|.....|.....|.....|.....|
332 roLAsnLeuThrPheMetSerGlnLeuAsnAsnProGlnAlaGlyGln 348
1036 TATATGGTAAATTTAATGTTACTTCACACCAAGTAGTCAACACTTC 1083
   ||| ||| .....|.....|.....|.....|.....|.....|.....|
349 TySerGlyIleLeuAsnValIleThrPheThrProSerSerSerLeu 364

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT: AAB45919
seq_documentation_block:
ID AAB45919 standard; Protein; 359 AA.
XX
AC AAB45919;
XX
DT 23-MAR-2001 (first entry)
XX
DE S. enterica serovar Typhi, tcfD fimbrial subunit protein.
XX
KW Fimbrial protein; saf; tcf; vaccine; gene therapy; immunization;
KW tcf insert; detection.
XX
OS Salmomella typhi.
XX
PM WO200073336-A1.
XX
PD 07-DEC-2000.
XX
PF 26-MAY-2000; 2000WO-SE01079.
XX
PR 28-MAY-1999; 99SE-0001961.
XX
PA (ACT1-) ACTIVE BIOTECH AB.
XX
PI Folkesson A, Normark S, Loefdaahl S;
XX
DR WPI: 2001-061512/07.
XX
DR N-PSDB: AAC82926.
XX
PT Fimbriae proteins of Salmomella enterica subspecies I bacteria, useful
PT for producing vaccines against the bacterial subspecies and for
PT detecting the bacteria
XX
PS Disclosure: Page 72-73; 77pp; English.
XX
CC This invention describes the novel proteins (saf and tcf) (I) encoded by
CC a DNA sequence of a gene encoding the precursor of the saf fimbriae unit
CC of Salmomella enterica subspecies I or a DNA sequence of the gene
CC encoding the tcf fimbriae unit of S. enterica subspecies I serovar Typhi
CC The products of the invention can be used as vaccines or for gene
CC therapy. Such vaccines are useful for protection against diseases caused
CC by S. enterica subspecies I or S. enterica subspecies I serovar Typhi.
CC The saf and tcf proteins from S. enterica subspecies I bacteria are
CC useful for active or passive immunization in mammals. The nucleotide
CC sequences are useful for constructing vectors for use as vaccines for
CC insertion into attenuated bacteria in constructing a recombinant viral
CC vaccine, or for direct inoculation of a nucleic acid vaccine. The protein
CC or antigenic fragments, nucleic acid sequences, and antibodies are useful
CC in molecular diagnostic assays for the detection of S. enterica
CC subspecies I.
XX
SQ Sequence 359 AA:
XX
alignment_scores:
Quality: 275.50 Length: 303
Ratio: 1.514 Gaps: 11
Percent Similarity: 60.066 Percent Identity: 29.703
alignment_block:

```


US-09-839-894-9 x AAB45919

Align seg 1/1 to: AAB45919 from: 1 to: 359

```
208 TTTTATGTTGTTCTCTGCACATACACTTAATGAGCATGTCACAAACAG 257
      : : : : : | | | | | : : : : : | | | | | : :
79 TrpValCysArgSerIserIsmrIrgsInuansInuGluGlyAlaCysGluGluTh 95
      : : : : : | | | | | : : : : : | | | | | : :
258 TGAAGATCCATACAGCTTCATCGCGTCACGGTGAAACAATTAACATTAC 307
      : : : : : | | | | | : : : : : | | | | | : :
95 rHisLeuValIrrTrpTyrAlaPheGlyAlaTyrSerIysIleAlaGlu 112
      : : : : : | | | | | : : : : : | | | | | : :
308 AATTTCAGGAAAAAGAACTTTAATAAAAAGAGACTACAAATTAAAGC 357
      : : : : : | | | | | : : : : : | | | | | : :
112 rXpHeArgGluInuIleSerHisAlaGluIleThrLeu..... 124
      : : : : : | | | | | : : : : : | | | | | : :
358 TATAAACAATTAATGTTCCAAAGGTGTAC.....TGCCATCCGGCCT 401
      : : : : : | | | | | : : : : : | | | | | : :
125 .....IleLeuLeuGlySerValAlaGspAlaCysTyrThrGlyVa 138
      : : : : : | | | | | : : : : : | | | | | : :
402 AACACTTAACATCAGCTCAATTTAACTGTAAATAAAACGGCGCTCA... 447
      : : : : : | | | | | : : : : : | | | | | : :
138 l.....IleasnMetasnAlaAlaIaCysG 147
      : : : : : | | | | | : : : : : | | | | | : :
448 ....GTCGCAAGTTATATTTATATATTCCTGCTGGCGCACTAAAAAT 492
      : : : : : | | | | | | | | | | | | | | | : :
147 InrrpGlyArgSerIleuLysLeuArgIleProSerGluInuLeuAlaIys 163
      : : : : : | | | | | | | | | | | | | | | : :
493 TTGCGCTTTGGTGCTGTCTGGGATCGCTACGTGAAAGTTA..AGAGTAA 539
      : : : : : | | | | | | | | | | | | | | | : :
164 lIleTrpThrSerIlyThrTyrLysAlaThrLeuValLeuAspTyrLeuG1 180
      : : : : : | | | | | | | | | | | | | | | : :
540 AAGACGATATAGAGAACCTTATGAACTTACACTTAATTAATCATTATTA 589
      : : : : : | | | | | | | | | | | | | | | : :
180 nTrpGlyGlyAspAspProLeuGlyThrSerThrIhrAspIleThrIleua 197
      : : : : : | | | | | | | | | | | | | | | : :
590 AATTAACTGAT.....AAGGCAATATTACATATGTTACCTCAGTTC 633
      : : : : : | | | | | | | | | | | | | | | : :
197 snValIhrAspHisPheAlaGluAsnAlaAlaIleTyrPheProGlnPhe 213
      : : : : : | | | | | | | | | | | | | | | : :
634 ...AAAGAGACGCGTCGCGTCGTCCTTACCTGCGCCCAACGTGGGG 680
      : : : : : | | | | | | | | | | | | | | | : :
214 GlyThrAlaThrProAlrGlyAlaSpLeuAsnLeuHisIAspMetAsnAlaSe 230
      : : : : : | | | | | | | | | | | | | | | : :
681 CACATATATTTGAGAAAGAAATTCGTGATGTATGCTTTATGATGATATA 730
      : : : : : | | | | | : : : : : | | | | | : :
230 rGlnMetSerGlyArgAlaAsnLeuAspMetCysLeuTyrAspGly..G 246
      : : : : : | | | | | : : : : : | | | | | : :
731 GTACTAACAAGACAGCTCTTTGAGATPAGATTTTCAGGATPAACATCTTAA 780
      : : : : : | | | | | : : : : : | | | | | : :
246 lValIlyAsnIaArgSerIleuGlnMetLysIleGluGlySerAsnIysSer 262
      : : : : : | | | | | : : : : : | | | | | : :
781 TCGATAGGGGAATTTATATTAAGGAAATAAATAATGACACCAAGAAGAT 830
      : : : : : | | | | | : : : : : | | | | | : :
263 GlyThrGlyPheGlnValIleLysSerAspSerAlaAspThr.....l 277
      : : : : : | | | | | : : : : : | | | | | : :
831 TGCATATACTTGTCTCACTCTCTGTGGCGGGTAAAGATTAATCTCCACAA 880
      : : : : : | | | | | : : : : : | | | | | : :
277 eaSpTyrAlaValSerMetAsnTyrGlyLysArgSerIleProValThrA 294
      : : : : : | | | | | : : : : : | | | | | : :
881 ATGGAACGTCATTAATAATGTCTGACGACAGCTTCTGTGAAACAACAGTGG 930
      : : : : : | | | | | : : : : : | | | | | : :
294 rGglIValIgluPheSerIleuAspAsnValaIAspAlaIaThrArg... 309
      : : : : : | | | | | : : : : : | | | | | : :
931 AATAGAATTAACAGCTGTGCACATGCCAGAAATCAGGTTCGGCTGTGG 980
      : : : : : | | | | | : : : : : | | | | | : :
310 .....ProValIleLeuProGlyGlnArgGlnAlaValaIArgCy 322
      : : : : : | | | | | : : : : : | | | | | : :
981 TTGCGCTGAGAGCTTTGCAATTTGGATGCAAAA.....GTGGAATAATC 1021
      : : : : : | | | | | : : : : : | | | | | : :
322 sValProValProLeuThrLeuThrThrGlnProPheAsnIleArgGlu 339
      : : : : : | | | | | : : : : : | | | | | : :
1022 CGAGCGCTGCACAAATATATGCGTAATATTAATGTTACTTCACACCAAGT 1071
```

[illegible]

513 etllelleatrgatrgatrhleuTYTYYPhepheanleulleleproc 530

841TTGCACTTCTTGGCGGTAAGTTAACTTCAACA 879

530 yValleullelleasermetalenleugilypherhileuproPropasp 546

880 AATGGAGCGTCAATTAAATTGCGAGCGAGCTCTCTGGA...ACAA 926

547 SerGlygluYsleuSerleqlyvaltrhleuenseerleuthra 563

927 CTGGAAATGAATTACAGCTGTCCACATGCCAAGAAATCACT.....GTTG 970

563 lPheleusnmetvalaIaIuItrhmetProalathSerAspAlaValp 580

971 CGGTG.....TTGTGTGG.....984

580 roleuTrPleIeargIleValPheleucystrPleuProTrIleuArg 596

985CCTGGACGT 993

597 MetSerArgProGlyArg 602

seq_name: /SIPSI/9cgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:AA47575

seq_documentation block: |

ID AAR47575 standard; Protein; 650 AA.

XX AAR47575; |

XX 19-JUL-1994 (first entry)

XX Alpha-agglutinin of *Saccharomyces cerevisiae*.

DE Immobilisation; enzyme; cell wall; alpha agglutinin; AGA 1; FLO 1;

XX Major cell wall protein; glycosyl-phosphatidyl-inositol;

KW anchoring protein; alpha factor; alpha-agglutinin; invertase;

KW inulinase; alpha-amylase; *Saccharomyces cerevisiae*;

XX enzymatic process; fermentation; biodegradation; catalysis.

XX *Saccharomyces cerevisiae*.

OS WO9401567-A.

PN 20-JAN-1994.

XX 07-JUL-1993; 93WO-EP01763.

XX 08-JUL-1992; 92EP-0202080.

PR 14-DEC-1992; 92EP-0203899.

XX (UNIL) UNILEVER NV.

XX (UNIL) UNILEVER PLC.

XX Kils FM, Schreuder MP, Toschka H, Verrips CT;

XX WPI: 1994-035071/04.

DR N-PSDB: AAO54012.

XX Immobilisation of enzymes to microbial cell wall - by prodn. of

PT fusion protein of enzyme linked to anchoring protein

PS Example 1; Page 32-39; 99pp; English.

XX The alpha-agglutinin is used in a method to immobilise enzymes to a

CC microbial cell wall. The coding sequence is used in the production

CC of a recombinant polynucleotide which comprises a structural gene

CC encoding a protein with catalytic activity and at least part of a

CC gene encoding at least the C-terminus of a protein capable of

CC anchoring in a eukaryotic or prokaryotic cell wall. The anchoring

CC fragment or protein is selected from alpha agglutinin, AGA 1, FLO 1

CC major cell wall protein of lower eukaryotes or a proteinase of

CC lactic acid bacteria. The recombinant polynucleotide preferably

CC also comprises a sequence encoding a signal peptide to ensure


```
806 AATAATGATGACACCAAGAAATTCATATCTTGTCTCTCTG 855
      |||||..... ||| :||| :|||
582 snueuthrlleysthrlsglueuthrleuthraspsnleuanlle 598
      :|||
856 GCGGCT.....AAAGTTTAACTCCACAATGGAACGTCATTAA 896
      :|||
599 SerGlyPheasnlysalagluilethrAlaLysAspnserspleuII 615
      :|||
897 TATTGTGACGCGCTCTCTGTGAAACAACTGGATAGATTACAGCTG 946
      |||||..... ||| :|||
615 eileeglyLysAlaSerSerAspnsersnAlaLysGlnlethrPhea 632
      :|||
947 TCACCATGCCAGAAATCAGTGTCCGGTGTGTGGCCCTGGACGTTTG 996
      :|||
632 splysValLysAspSerLysIleSer.....AlaGlyAsnHisAsnVal 646
      :|||
997 CAATTGGATGCAAAAGTGGAAATCCCGAGGCT..... 1029
      :|||
647 ThrLeuAsnSerLysValGluThrSerAsnSerAspGlySerThrGlyAs 663
      :|||
1030 .GGACAATATATGGTAAATTAATGTTACTTTCACACCAAGATGCAAA 1078
      |||||..... ||| :|||
663 nglySerAspAspAsnAsnIleGlyLeuThrIleSerAlaLysAspValT 680
      :|||
1079 CACTC 1083
      :|||
680 hrVal 681
```

seq_name: /SIDSL/gcgdata/hoid-geneseq/geneseq-emb1/AA1993.DAT:AA41732

seq_documentation_block:

ID AAR41732 standard; Protein; 1529 AA.

AC AAR41732;

DT 26-APR-1994 (first entry)

DE High molecular weight protein 4 (HMW4).

KW HMW: high molecular weight protein; virus; vaccine; influenza;

KN epitope; immunity; haemophilus influenzae.

OS Haemophilus influenzae.

PN WO9319090-A.

PD 30-SEP-1993.

PF 16-MAR-1993; 93WO-US02166.

PR 16-MAR-1992; 92GB-0005704.

PA (BARE/) BARENKAMP S J.

PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

PI Barenkamp SJ;

XX WPI: 1993-320683/40.

DR N-PSDB; AAO49511.

PT High molecular weight surface proteins - of non-typeable
haemophilus which exhibit immunogenic properties

PS Claim 6; Figure 10; 100pp; English.

CC The isolation and purification of the high molecular weight protein
enables the identification of the major protective epitopes of the
protein by conventional epitope mapping. These epitopes can then be
synthesised using standard techniques and incorporated into fully
synthetic or recombinant vaccines.

SO Sequence 1529 AA;

alignment_scores:

Quality:	110.50	Length:	348
Ratio:	0.594	Gaps:	18
Percent Similarity:	53.448	Percent Identity:	20.690

alignment_block:

US-09-839-894-9 x AAR41732

Align seg 1/1 to: AAR41732 from: 1 to: 1529

```
142 CATATATATATTAATTAACATATTACAGATACAGTGAAGTCATCTCT 191
      :|||
744 TyrAsnGluTyrSerLysHis.....AlaIleAsnSerSerHisAsnIle 758
      :|||
192 GTATGATAGATGACTTTTATGTTTGCTCTCTCATACACTTAATG 241
      :|||
758 u.....ThrIleLeuGlyAsnValThrLeuGlyG 769
      :|||
242 GACCATGTCCAAACAGTGAAGATCTAGAGTTCA..... 276
      :|||
769 Ly.....GluAsnSerSerSerSerIleThrGlyAsnIle 780
      :|||
277 TCGTCAGCGGTGAACAATATATACATTACATTT..... 312
      :|||
781 AsnIleThrAsnLysAlaAsnValThrLeuGlnAlaAspThrSerAsnSe 797
      :|||
313 .....ACGGAAGAAAGAGTTTAATTAAGAGAGCTCAAAATTA 352
      :|||
797 rAsnThrGlyLeuLysLysArgThrLeuThrLeuGlyAsnIleSerValG 814
      :|||
353 AAGCTATATAACAATATTGTTTCAAAAGTGTAACTGCCATCCGGCTTA 402
      :|||
814 LucIlyAsnLeuSerLeuThrGlyAlaAsnAlaAsnIleValGlyAsnLeu 830
      :|||
403 ACACTTAACTCAGCTCATTTTAACTGTAATTAAGAACCGGCTTCAGGTGC 452
      :|||
831 SerIle.....AlaGluAspSerThrPheLysGlyAlaAlaSer..... 843
      :|||
453 AACTTATATATTATATATTCCTCTCGCGCAACTAAAAAATTTCCTTTTG 502
      :|||
844 .....AspAsnLeuAsnIleThr 849
      :|||
503 GTGCTATCTGG.....GATGCTACTCTGTAAGTTAAGATAAAGACGCA 546
      :|||
849 hrcGlyThrPheThrAsnAsnGlyThrAlaAsnIleAsnIleLysGlnGly 865
      :|||
547 TATAGTGACACTATGGAACCTTACCT.....ATAAATATCAC 584
      :|||
866 ValValLysLeuGlnGlyAspIleIleAsnLysGlyLeuAsnIleThr 882
      :|||
585 TATTAATTTACT..... 597
      :|||
882 rThrAsnAlaSerGlyThrGlnLysThrIleIleAsnGlyAsnIleThrA 899
      :|||
598 ..GATAAGGAATATTACAGATATGCTTACCTCAGTTCAAAAGTCAGCT 645
      :|||
899 snGluLysGlyAspLeuAsnIle.....LysAsnIleLysAlaAspAla 913
      :|||
646 CCGGTCGATCTT.....AACTTGCCTCCAACTGTGGGGGACATATAT 689
      :|||
914 GluIleGlnIleGlyLysenIleSerGlnLysGlnGlyAsnLeuThrII 930
      :|||
690 TGAAGAAATTCGTGATATATGCTTTATATGATATAGATACATACAA 739
      :|||
930 eSerSerAspLysValAsnIle.....ThrAsnG 940
      :|||
740 GCGAGCTTTGGAGATACATTCAG.....GATAACAATCTCT 777
      :|||
940 InlethrIleLysAlaGlyValGlnGlyGlyArgSerAspSerSerGlu 956
      :|||
778 AAATCTGATGGAAATTTATCTAAGCAAAATTAATGATGACACCAAGA 827
      :|||
```

```

957 AlaGluAsnAlaAsnLeuThrIleGln.....ThrLysG1 968
828 AATGCAATACCTTGTCTCTCTGCGGCT.....AAAGTT 868
      ::::: ::::: ::::: ::::: :::::
968 uLeuLysLeuAlaGlyAspLeuAsnIleSerGlyPheAsnLysAlaIu1 985
      ::::: ::::: ::::: ::::: :::::
869 TAACTCAACAAATGACGATCATTAATATGCTGACGAGCTTCTCTG 918
      ::::: ::::: ::::: ::::: :::::
985 IeThrAlaLysAsnGlySerAspLeuThrIleGlyAsnAlaSerGlyGly 1001
      ::::: ::::: ::::: ::::: :::::
919 GAACAACAATGGAATAGATTACAGCTGCACCATGCCAGAAATCAGTGT 968
      ::::: ::::: ::::: ::::: :::::
1002 AsnAlaAspAlaLysLys.....ValThrPheAspLysValLysAs 1015
      ::::: ::::: ::::: ::::: :::::
969 TCCGCTGTGTGTTGGCTTGA...CGTTGCAATGATGCAAAATGCG 1015
      ::::: ::::: ::::: ::::: :::::
1015 pSerLysIleSerThrAspLysAsnValThrLeuAsnSerIleVal1 1032
      ::::: ::::: ::::: ::::: :::::
1016 AAATCCCGAGCTGCACAAATATGATGTAATTAATGTTACT 1059
      ::::: ::::: ::::: ::::: :::::
1032 ystHrSerAsnGlySerSerAsnAlaGlyAsnAspAsnSerThr 1046
      ::::: ::::: ::::: ::::: :::::
seq_name: /SIDSI/gcgsdata/hold-geneseq/geneseqp-emb1/AA1997.DAT:AAW30292
seq_documentation_block:
ID AAW30292 standard: Protein; 1601 AA.
XX
AC AAW30292:
XX
DT 14-APR-1998 (first entry)
XX
DE Non-typeable Haemophilus high mol.wt. surface protein HMW4.
XX
KW Non-typeable Haemophilus; high molecular weight surface protein;
HMW4; immunogen; vaccine; otitis media.
XX
OS Haemophilus influenzae strain 5.
XX
FH Key Location/Qualifiers
FT Misc-difference 372 /note= "encoded by TCR"
FT Misc-difference 400 /note= "encoded by AAT"
FT Misc-difference 400 /note= "encoded by AAT"
XX
PN W09736914-A1.
XX
PD 09-OCT-1997.
XX
PF 01-APR-1997; 97WO-US04707.
XX
PR 01-APR-1996; 96US-0617697.
XX
PA (BARE/) BARENKAMP S J.
XX
PI Barenkamp SJ;
XX
DR WPI: 1997-503038/46.
DR N-PSDB: AAT90993.
XX
PT High molecular weight proteins of non-typeable Haemophilus
PT Influenzae - useful for vaccine production
XX
PS Claim 1: Page 97-102; 183pp; English.
XX
CC This protein comprises the high molecular weight surface protein
CC HMW4 (123 kDa) of non-typeable Haemophilus influenzae strain 5 that
CC has the immunological ability to protect against disease caused by
CC a non-typeable Haemophilus strain and is characterised by at least
CC one surface-exposed B-cell epitope that is recognised by monoclonal
CC antibody AD6. The HMW4 amino acid sequence was deduced from an
CC isolated HMW4 gene (see AAT90993). HMW1 (see AAW30293), HMW2 (see
CC AAW30294) and HMW3 (see AAW30291) have also been identified. A
CC conjugate comprising HMW4 linked to an antigen, hapten or

```

CC polysaccharide, and a synthetic peptide of 6-150 amino acids
 CC corresponding to at least protective epitope of HMW4 are also
 CC claimed. HMW proteins, conjugates and peptides can be used in
 CC vaccines, as immunogens for preparation of antibodies and as
 CC antigens for detection of these antibodies.

SO Sequence 1601 AA;

alignment_scores: quality: 110.50 length: 348
 ratio: 0.594 gaps: 18
 Percent Similarity: 53.448 Percent Identity: 20.690

alignment_block:
 US-09-839-894-9 x AAW30292 ..

Align seq 1/1 to: AAW30292 from: 1 to: 1601

```

142 CATATATATATTAATTAACATATTACAGCATGCAAGTCACTACTCT 191
      ::::: ::::: ::::: ::::: :::::
816 TyrAsnGlyThrSerLysHis.....AlaIleAsnSerSerHisAsn1e 830
      ::::: ::::: ::::: ::::: :::::
192 GTATGATAGCATGACTTTTATGTTGTCTCTCACAATACACTAATG 241
      ::::: ::::: ::::: ::::: :::::
830 u.....ThrIleLeuGlyGlyAsnValThrLeuGlyG 841
      ::::: ::::: ::::: ::::: :::::
242 GAGCATGTCCACACGATGAGATCCCTAGCAGTCA..... 276
      ::::: ::::: ::::: ::::: :::::
841 Ly.....GluAsnSerSerSerSerIleThrGlyAsn1e 852
      ::::: ::::: ::::: ::::: :::::
277 TCGGTACGCGGTGAACAATATATACATTCAATT..... 312
      ::::: ::::: ::::: ::::: :::::
853 AsnIleThrAsnLysAlaAsnValThrLeuGlnAlaAspThrSerAsnSe 869
      ::::: ::::: ::::: ::::: :::::
313 .....ACGGAAGAAAGAGATTAAATTAATTAAGAGAGCTACAAATTA 352
      ::::: ::::: ::::: ::::: :::::
869 rAsnThrGlyLeuLysLysArgThrLeuThrLeuGlyAsnIleSerValG 886
      ::::: ::::: ::::: ::::: :::::
886 IuGlyAsnLeuSerLeuThrGlyAlaAsnAlaAsnIleValGlyAsnLeu 902
      ::::: ::::: ::::: ::::: :::::
403 ACACCTTAACCTCAGCTCACTTTTAAGTAATTAAGCGGCTTCAAGTGC 452
      ::::: ::::: ::::: ::::: :::::
903 SerIle.....AlaGluAspSerThrPheLysGlyGluAlaSer..... 915
      ::::: ::::: ::::: ::::: :::::
453 AAGTTATATTTATATATTCCTGCTGGCAACTAAATAATTTGCTTTTG 502
      ::::: ::::: ::::: ::::: :::::
916 .....AspAsnLeuAsnIleT 921
      ::::: ::::: ::::: ::::: :::::
503 GTGCTATCTGG.....GATGCTACTCTGAAGTTAAGATTAAGAGAGA 546
      ::::: ::::: ::::: ::::: :::::
921 hrGlyThrPheThrAsnAsnGlyThrAlaAsnIleAsnIleLysGlnGly 937
      ::::: ::::: ::::: ::::: :::::
547 TATAGTGAGACTATGGAAGCTTACACT.....ATAATATATCAC 584
      ::::: ::::: ::::: ::::: :::::
938 ValValLysLeuGlnGlyAspIleIleAsnLysGlyLeuAsnIleThr 954
      ::::: ::::: ::::: ::::: :::::
585 TATTAATTAAT..... 597
      ::::: ::::: ::::: ::::: :::::
954 rThrAsnAlaSerGlyThrGlnLysThrIleIleAsnGlyAsnIleThrA 971
      ::::: ::::: ::::: ::::: :::::
598 ..GATAAGGGAATATTCAGATATGTTACCTCAGTTCAAAAGTACGCT 645
      ::::: ::::: ::::: ::::: :::::
971 snGluLysGlyAspLeuAsn1e.....LysAsnIleLysAlaAspAla 985
      ::::: ::::: ::::: ::::: :::::
646 CGCGTCATCT.....AACTTCGCTCAACTGCTGGGGGCAATATAT 689
      ::::: ::::: ::::: ::::: :::::
986 GluIleGlnIleGlyGlyAsnIleSerGlnLysGlnLysLeuThrI1 1002
      ::::: ::::: ::::: ::::: :::::
690 TGAAGAATCTCTGATATGCTCTTTATGATGATATATGATCTATACA 739
      ::::: ::::: ::::: ::::: :::::

```


[illegible]


```

151 .....TTAATAACAAATTACAGCATACAGTGAAGTCACTACTCTCTA 194
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
309 GlyAspValAsnPheHisPheAsnHisLaserSerAsnTyrGlnThrTyr 325
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
195 TGATAGCATGACTTTTATGTTGCTCTTCACAAAT..... 231
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
325 rGly.....ValIleIleGluSerGlnAsnPheSerAlaSerG 338
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
231 ..... 231
338 LysGlySerSerLeuLysPheLysSerGluGlySerThrHisAlaAlaPhe 354
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
232 .....ACACT 236
355 ThrIleLysAsnAspLeuIleLeuAsnAlaThrGlyLysAsnIleSerIle 371
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
237 TAATGAGCATGTCACACCGAGAAATCTAGAGTTCATCGGTGACG 286
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
371 uAsnGlnValAlaGlyIleAspSerAsnLeuLysSerLeuIleAla 388
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
287 GTGAACA.....AATATACATTACAAATTACGGA 318
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
388 snLysAsnIleThrPheGluGlyLysAsnIleThrLeuAlaAlaAspLys 404
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
319 AAAAGAAGTTAATATAAAGAGACTACAAATTAAGGCTTAACAATT 368
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
405 LysProIleGluIleLysGlyLysAsnIleThrValLysGluGlyAlaAsnVa 421
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
369 ATTGTCAAAGTGTAACTGC..... 390
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
421 LThrLeuArgSerLysAsnTyrGlyAsnAspLysSerAlaLeuSerIleA 438
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
391 .....CCATCGCGCTACACCTTAACCTACGTCATTTT 423
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
438 rGlyAsnValThrAsnLysGlyAsnLeuThrValThrGlySerAlaIle 454
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
424 AACGTATATAAANC..GCCGCTTCAGCTGCACAGTTATATTATAT 470
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
455 AsnIleGluLysAsnLeuThrValGlnLysSerLysPheLeuAlaAs 471
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
471 TCCTGCTGCGCACTAAAAAATTGCTCTTT.....GGTGTATCTGG 514
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
471 nPro.....AsnTyrSerPheAsnValSerGlyLeuPheA 483
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
515 ATGCTACTCTGAAGTTAAGTAAAA.....AGACGATAT 549
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
483 spAsnGlnGlyLysSerAsnIleSerIleAlaLysGlyAlaIlePhe 499
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
550 AGTGAGACTTGAAGCTTACACTATAATATCATTAATAATTACTGA 599
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
500 LysAspIleGluAsnThrGlySerLeuAsnIleThrThrLysSerAspSe 516
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
600 T.....AAGGAAATATTCAGATATGTTACTCTAGT 631
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
516 rAsnHisHisThrIleIleLysGlyAsnIle..... 526
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
632 TCAAAAGTACGCTCGCTCGATCTTAACCTGCTCACTCACTGGGGGC 681
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
527 .....ThrAsnArgLysGlyAspLeuAsnIleThrAsnAsnGlyAspAsn 541
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
682 ACA.....TATATTGGAAGAATTCGTGTATATGCTTTTATGATGG 725
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
542 ThrGlnIleGlnIleGlyGlyAsn.....IleSerGlnLysGluG 555
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
726 ATATAGTACTACAGACAGCTCT.....TTGAGA 754
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
555 yAsnLeuThrIleSerSerAspLysValAsnIleThrGlnArgIleTrp 572
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
755 TAAGATTTCAGGATACAAATCCTAAATCTGAT...GGGAAATTTTATCTA 801
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
572 LelysAlaGlyValAlaAsnGlyAspAsnSerAspSerAsnGlnAlaThrSer 588
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

802 AGGAATAATATGATGACACCAAGAAATGCAATFACTTTGTCACTTCT 851
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
589 AlaAsnLeuThrIleLysThrLysGluLeuLysLeuThrAsnAspLeuAs 605
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
852 CTGGCGGGT.....AAAAGTTTAACCTCCAAACAATGGAAGCTCAT 892
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
605 nIleSerLysPheAsnLysAlaGluIleThrAlaLysAspAsnSerAsnL 622
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
893 TAAATATTGCTGACGACGCTTCTCTG...GAACAACAATGCAATGAAAT 939
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
622 eutThrIleGlyAspAsnSerAspAlaGlyAsnThrAspAlaLysLysVal 638
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
940 ACAGCTGCACCAATGCCAGAAATCAGTGTCCGGGTGCTGTGGCCCGG 989
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
639 ThrPheSerAsnValLysAspSerLysIleSerAlaSer.....AspHI 653
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
990 ACGTTTGCAATTGGATGCAAAAGTGGA.....AATCCGAGGCTG 1030
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
653 sAsnValThrLeuAsnSerLysValGluThrSerGlyAspThrAspSerT 670
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1031 GACAAATATATGGTAATATATGTTACTTTCACACCAAGTAGTCAACA 1080
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
670 hrGluAspGlyLysAsnAsnAsnThrGlyLeuThrIleThrAlaLysAsn 686
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1081 CTC 1083
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
687 Val 687
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:AA01828
seq_documentation_block:
ID AA01828 standard; protein: 1228 AA.
AC AA01828;
DT 11-SEP-2000 (first entry)
DE Haemophilus influenzae strain KI full-length HMW1A protein. SEQ ID NO:34.
KW HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW recombinant production; Escherichia coli; antibacterial; vaccine;
KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW detection; diagnosis.
OS Haemophilus influenzae strain KI.
FH Key Location/Qualifiers
FT Misc-difference 313 /note="Encoded by GG"
FT WO200020609-A2.
PD 13-APR-2000.
PF 07-OCT-1999; 99WO-CA00938.
PR 07-OCT-1998; 98US-0167568.
PR 08-DEC-1998; 98US-0206942.
PA (CONN-) CONNACHT LAB LTD.
XX Iloosmore SM, Yang Y, Klein MH;
XX WPI: 2000-303789/26.
XX DR N-PSDB; AAA52179.
XX PT Nucleic acid molecule for producing recombinant high molecular weight
XX proteins of Haemophilus which are used as a vaccine to provide
XX protection against Haemophilus induced diseases in humans -
XX Claim 12; Fig 20A-R; 307pp; English.
XX

```


922 ACAACATCGGAT.....AGAAITTCACCTGTCCACCATCCAGAATCAG 965
||| ||| : : : : : ||| |||
1143 GlysAlaIleAsnAlaAlaAsnValThrLeuAsnTrpThrGlyThrLeuTh 1159
::: :::
966 TGT.....CCGGTGTGTGTGGCGCTGCACCTTGCAATTG.. 1002
::: :::
1159 rThrThrGlyAspSerLysIleAsnAlaThrSerGlyThrLeuThrIleA 1176
::: :::
1003GANGCAAACACTGCAAAATCCCAGACCTGGA..... 1032
||||| ||| : : : : : ||| |||
1176 snAlaLysAspAlaLysLeuAspGlyAlaAlaSerGlyAspArgThrVal 1192
::: :::
1033CATATATGGGTATAATTATGTACTTTACACACCAG 1070
::: ||| : : : : : ||| |||
1193 ValAsnAlaThrAsnAlaSerGlySerGlyAsnValThrAlaLysThrSe 1209
1071 TAGT 1074
||| |||
1209 rSer 1210


```

151 TTAATACCATATATACAGCATACAGTGAAGTCACTATCTATGATAG 200
|||||.....|
51 LeuAsnSerThrLeuThrAlaTyrAsnGlySerHisHisLeuTyrAsp 67
|||||.....|
201 GATGACTTTTATGTTTGTCTTCACAAATACACTTAATGAGCATGTC 250
|||||.....|
67 gmetSerPheLeuCysLeuSerSerGlnSerThrLeuAsnGlyAlaCysP 84
|||||.....|
251 CAACGAGTGAATCCTAGAGTTCATCGGTCAGCGGTGAACCAATATA 300
|||||.....|
84 roSerSerAspAlaProGlyThrAlaThrIleAspGlyLysTrpAsnIle 100
|||||.....|
301 ACATTACAAATTAAGGAAAAAGAACTTAATTAAGAGAGCTACAAAT 350
|||||.....|
101 ThrLeuGlnPheThrGlnLysArgSerLeuIleLysArgGlnLeuGlnI 117
|||||.....|
351 TAAAGGCTTAACAACTATATGTTCAAAAGTGAATCTGCCATCGGCC 400
|||||.....|
117 eLysGlyTyrLysGlnPheLeuPheLysAsnAlaAsnCysProSerLys 134
|||||.....|
401 TTAACCTTAACCTCAGCTCATTTTAACTTAATTAAGGAGGCTTACG 450
|||||.....|
134 euAlaLeuAsnSerSerHisPheGlnCysAsnArgGlnAlaSerGly 150
|||||.....|
451 GCAAGTTTATTTATATATTCCTGCTGGCGAAGCTAAAAAATTTGCC 500
|||||.....|
151 AlaThrLeuSerLeuTyrIleProAlaGlyGlnLeuAsnLysLeuPro 167
|||||.....|
501 TGGTGTATCTGGAGTCTACTGTAAGTTAAGAGTAAAAAGCATATA 550
|||||.....|
167 eGlyGlyValTyrPasnAlaValLeuLysLeuAsnValLysArgArgTyr 184
|||||.....|
551 GTGAGACCTTGAGACTTACATATTAATTAATTAATTAATTAATGAT 600
|||||.....|
184 sphThrThrTyrGlyThrTyrThrIleAsnIleThrValAsnLeuThr 200
|||||.....|
601 AAGGAAATATTCAGATATGCTTACCTCAGTTCAAAAGTCAGCTGCGT 650
|||||.....|
201 LysGlyAsnIleGlnIleThrPleuProGlnPheLysSerAsnAlaArg 217
|||||.....|
651 CGATCTTAATCTGCGCAACTGCTGGGGGACATATATTGGAAGAAT 700
|||||.....|
217 LaspLeuAsnLeuArgProThrGlyGlyThrTyrIleGlyArgAsn 224
|||||.....|
701 CTGTGATATGCTTTTATGATGATATAGTACTAAGCAGCAGCTCTTG 750
|||||.....|
234 eValAlaPheTyrSerPheTyrAspGlyTyrSerThrAsnSerSerLeu 250
|||||.....|
751 GAGATTAAGATTCAGATACCAATCCTAAATCGATGGAAATTTATCT 800
|||||.....|
251 GlnIleArgPheGlnAspAsnSerLysSerAspGlyLysPheTyrIle 267
|||||.....|
801 AAGGAAATTAATGATGACACCAAGAAATTCATATCTTGTCTTC 850
|||||.....|
267 uLysLysIleAsnAspSerLysGlnLeuValTyrThrLeuSerLeu 284
|||||.....|
851 TCTTGCGGGTAAAGTTTAACCTCAACCAAAATGAGAGCTCAATTA 900
|||||.....|
284 euLeuAlaLeuLysAsnLeuThrProThrAsnGlyAlaLeuAsnIle 300
|||||.....|
901 GGTGAGCAGACTTCTCTGAAACCAATCGAATAGATTAACAGCTGTC 950
|||||.....|
301 ...AsnThrAlaSerLeuGlnThrAsnTrpAsnArgIleThrAlaVal 316
|||||.....|
951 CATGCCAAGAAATCAGTGTCCGGTGTGTTGGCTGAGCAGTTGCCAAT 1000
|||||.....|
316 rMetProGlnIleSerValProValLeuCysTrpProGlyArgLeuGln 333
|||||.....|
1001 TGGATGCAAAAGTGAAGAAATCCGAGCTGGGACATATATGTTATAT 1050
|||||.....|
333 euAspAlaLysValLysAsnProGlnAlaGlyLysTyrSerGlyIleLeu 349
|||||.....|
1051 AATGTTACTTTCACACCAAGTAGTCAACACATC 1083

```

```

|||||.....|
350 AsnValThrPheThrProSerSerSerLeu 360
seq_name: /cgn2_6/ptodata/2/laa/5B_COMB.pep:US-08-483-101-16
seq_documentation_block:
; Sequence 16, Application US/08483101
; Patent No. 5932715
; GENERAL INFORMATION:
; APPLICANT: Scott, June R.
; APPLICANT: Froehlich, Barbara
; APPLICANT: Caron, Judy
; TITLE OF INVENTION: CS2 Proteins and Coding Sequences
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee and Winner, P.C.
; STREET: 5370 Manhattan Circle, suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,101
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Feiber, Donna M.
; REGISTRATION NUMBER: 33878
; REFERENCE/DOCKET NUMBER: 6-95
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 363 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; US-08-483-101-16

alignment_scores:
Quality: 988.50 Length: 371
Ratio: 3.317 Gaps: 11
Percent Similarity: 80.323 Percent Identity: 54.987

alignment_block:
US-09-839-894-9 x US-08-483-101-16 ..
Align seg 1/1 to: US-08-483-101-16 from: 1 to: 363

7 AGATTATTTATTTATTTTACATGTTTTCCTGAGTCTTTTACAT 56
||| :|||
||| :|||
2 LysIlePhePheLeu.....SerIlePheSer 14
||| :|||
57 TCGTATGAGGAGATAAATTCGCGAGATGA.....AGATTAACA 100
||| :|||
14 aValValSerAlaGlyArgTyrProGlnThrValGlnAsnLeuThr 31
||| :|||
101 ATATTTT...GGCCGCGT...GACAGAGCAATCTCCCAACAT 144
||| :|||
31 ySerPheGlnAlaProAlaGlyLeuAspArgSerValGlnSerProIleTyr 47
||| :|||
145 AATATATTAATTAACCATATTAACAGCATACAGTGAAGTCACTCTGTA 194
|||||.....|
48 AsnIlePheThrAsnIleValAlaGlyTyrSerLeuSerHisSerLeuTyr 64

```

```

195 TGATAGATGACCTTTTATGTTGCTCTCACAATACCTTAATGAG 244
64 RASPRGLLEVALPHEUCYSTHRSESSERSESPROVALASNGLYA 81
245 CATGTCCACACGAGTGAATCTAGACAGTTCATCGTCAGC...GGTAA 291
81 IACYSPTHTFLE.....GLYTHRSERGLYVALGINTTYRGLYTHR 94
292 ACAATATACATTACATTTAGCGAAAAAGATTAAATAAAGAGA 341
95 ThrThrIleThrLeuGlnPheThrGluLysArgSerLeuIleLysArgAs 111
342 GCTACAAATTAAGGCTATAACATATGCTCAAAAAGTTTAACTCC 391
111 ntleasneulagllyasnlylsprollletpgluasnlinsercysa 128
392 CATCCGCG.....CTAACCTTAACCTCAGCTCATTTAACTGAATAA 435
128 sPheSerAsnLeuMetValLeuAsnSerLysSerTrpSerCysGlyAla 144
436 AACCGC...GCTCAGGTGCAAGTTTATATTTATATATTCCTGCGCA 482
145 HisGlyAsnAlaAsnGlyThrLeuLeuAsnLeuTyrlleProAlaGlyG 161
483 ACTAAAAATTTGGCTTTGGTGTATCTGGGAGCTACTCTGAAGTTAA 532
161 uilasnlylsprollletpgluaglthrleuileua 178
533 GAGTAAAGACGATATAGTGAGAC.....TATGCAACTTAC 570
178 rgleuser...ArgTyrGlyGluValSerSerThrHisTyrGlyAsnTyr 193
571 ACTATAATATCTATTTAAATTAAGTGAATTAAGCAATATTCAGATAG 620
194 ThrValAsnIleThrValAspLeuThrAspLysGlyAsnIleGlnValTr 210
621 GTTACCTCAGTCAAAAGTGACGCTCGCTGATCTTAACCTGCTCCAA 670
210 PLeuProGlyPheHisSerAsnProArgValAspLeuAsnLeuArgPro 227
671 CTGCTGGGGCAGCATATATTTGAGAAATTCGTGATATGCTTTTAT 720
227 leglyAsnTyrLysTyrSerLysSerAsnSerLeuAspMetCysPheTyr 243
721 GATGATATAGTACTTAACAGCAGCTTTTGAGATTAAGATTTCAGATTA 770
244 AspGlyTyrSerThrAsnSerAspSerMetValIleLysPheGlnAspAs 260
771 CAATCCTTAATCTGATGGAAATTTTATCTAAGAAAAATAATGATGACA 820
260 pAsnProThrAsnSerSerGluTyrAsnLeuTyrLysIleGly...glYt 276
821 CCAAGAAATGATCATATCTTTGACCTTCTTGGCGGAGTAAAGTTTA 870
276 hrGluLysLeuProTyrAlaValSerLeuLeuMetGlyGluLysIlePhe 292
871 ACTCCAAACAAATGAGACGCTCATTAATAATTCGTGACGAGCTTCTGGA 920
293 TyrProValAsnGlyLysSerPheThrIleAsnSpSerSerValLeuG 309
921 AACCAACTGGAATAGATTAAGCTGTACCATCCAGAAATCACTGTC 970
309 uThAsnTrpAsnArgValThrAlaValAlaMetProGluValAsnValP 326
971 CGGTGTGTGTGGCTGACGCTTGCATTAATGATGCAAAAGTGAATAT 1020
326 rovalleuLysTyrProAlaArgLeuLeuAsnAlaAspValAsnAsn 342
1021 CCCGAGGCTGACAAATATATGGTAAATTAATTAATCTTACACACCAAG 1070
343 ProGluAlaGlyGlnTyrMetGlyAsnIleLysIleThrPheThrProse 359

```

```

1071 TAGTCAACACATC 1083
359 rSerGlnThrLeu 363

```

```
seq_name: /cgn2_6/prodata/2/laa/5B_COMB.pep:US-08-483-101-5
```

```
seq_documentation_block:
```

```

; Sequence 5, Application US/08483101
; Patent No. 5932715
; GENERAL INFORMATION:
; APPLICANT: Scott, June R.
; APPLICANT: Froehlich, Barbara
; APPLICANT: Caron, Judy
; TITLE OF INVENTION: CS2 Proteins and Coding Sequences
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee and Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,101
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33878
; REFERENCE/DOCKET NUMBER: 6-95
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 364 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-483-101-5

```

```
alignment_scores:
```

```

Quality: 927.50 Length: 366
Ratio: 3.198 Gaps: 3
Percent Similarity: 79.235 Percent Identity: 48.087

```

```
alignment_block:
```

```
US-09-839-894-9 x US-08-483-101-5 ..
```

```
Align seg 1/1 to: US-08-483-101-5 from: 1 to: 364
```

```

1 ATGATTAATGATTTATTTATTTACATTTGTTTCTCTCAGTACTTT 50
1 LeuLysValIlePheValLeuSerMetPheLeuLysSerGlnValTyr 17
51 TACATTTCTGTATCGCAGATAAATTCGCGAGATCAAAAGCATATA 100
17 rGlyLysSerTrpHisThrAsnValGluAlaGlySerIleAsnLysThr 34
101 ATATTTTGGCCGGGTGACGAGAAC.....GAATCTTCCCAACAT 144
34 IuSerIleGlyProIleAspArgSerAlaAlaIleSerTyrProAlaHis 50
145 AATATATTAAATTAACATATTACAGCATTAAGTGAAGTCACTAGTA 194
51 TyrIlePheHisGlnHisValAlaGlyTyrAsnLysAspHisSerLeuPh 67

```

```

195 TGATAGAGTACTTTTATGTTGCTCTCAACAATACACTTAATGAG 244
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
67 eAspArgMetThrPheLeuCySmetSerSerThrAspAlaSerLysGly 84
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
245 CATGTCCACCAAGAGAGATCTAGCAGTTCATCGGTGACGGTGAACA 294
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84 lacysProThrGlyGluAsnSerLysSerSer.....GlnGlyGluThr 98
295 AATATACATTACCAATTTACGAAAAAAGAGTTAATAAAAAAGAGCT 344
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
99 AsnIleLysLeuIlePheThrGluLysLysSerLeuAlaArgLysThrLe 115
345 ACAATTAAGAGCTATAACAAATTATTGTTCAAAAGTTAACTGC.... 390
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
115 uAsnLeuLysGlyTyrLysArgPheLeuTyrGluSerAspArgCysIleH 132
391 ....CGATCGGCGCTAACACTTACCTACAGCTCACTTTAACTGAATAA 435
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
132 lStyValAlaPlyMetAsnLeuAsnSerHisThrValLysCysValGly 148
436 AACGGCGCTCAGGTCGAGTTATATATATATATTCCTGCTGCGCACT 485
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
149 SerPheThrArgGlyValAspPheThrLeuTyrIleProGlnGlyGluI 165
486 AAAAAAATTTGCTTTGCTGATCTGGAAGTCTGAACTGAAGTTAAGAG 535
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
165 eAspGlyLeuLeuThrGlyGlyIleThrPgluAlaThrLeuGluLeuArg 182
536 TAAAAAGACATATAGTACAGCACTATGGAACCTTACACTATAAATATCACT 585
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
182 alLysAlaGlnHisTyrAspTyrAsnHisGlyThrTyrLysValAsnIleThr 198
586 ATTAATATCTGATAGGAAATATTCAGATGATGTTACCTGACGTCAAC 635
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
199 ValAspLeuThrAspLysGlyAsnIleGlnValThrPheProLysPheH 215
636 AAGTGCAGCTCGCGTCGATCTTAAGTGCCTCAACTGATGGGGGCACAT 685
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
215 sSerAspProArgIleAspLeuAsnLeuArgProGluGlyAsnGlyLysT 232
686 ATATGGAAGAAATTCGTGATATGCTTTTATGATGATATATGATCT 735
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
232 ySerGlySerAsnValLeuGluMetCysLeuTyrAspGlyTyrSerThr 248
736 AACAGCAGCTCTTGGAGATAGATTAAGATTAACCAATCTTAATCTGA 785
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
249 HisSerGlnSerIleGluMetArgPheGlnAspSerGlnThrGlyAs 265
786 TGGGAAATTTTATCTAAGAAATTAATGATGACCAACCAAGAAATTCAT 835
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
265 naSngLutTyAsnLeuIleLysThrGlyIuProLeuLysLysLeuProt 282
836 ATACTTTGTACTTCTTGGCGGTAAAGTTAACTCCAACAATATGCA 885
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
282 yTrLysLeuSerLeuLeuGlyGlyArgGluPheTyrProAsnAsnGly 298
886 ACCTCAATTAATATGCTGACGAGCTTCTCGAAAAACAACAGGAATAG 935
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
299 GluAlaPheThrIleAsnAspThrSerSerLeuPheIleAsnThrPasnAr 315
936 AATTACAGCTGTGACCATGCCAGAAATCACTGTTCCGTTGTTGTTGGC 985
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
315 gIleLysSerValSerLeuProGlnIleSerIleProValLeuLysTrpP 332
986 CTGGACGTTTGCATTTGATGCAAAAGTGAATAATCCGAGGCTGACAA 1035
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
332 roAlaAsnLeuThrPheMetSerGluLeuAsnAsnProGluAlaGlyL 348
1036 TATATGGGTATATATTAATTAATTAATTAATTAATTAATTAATTAAT 1083
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
349 TyrSerGlyIleLeuAsnValThrPheThrProSerSerSerSerLeu 364

```

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-617-697-10

```

seq_documentation_block:
: Sequence 10, Application US/08617697
: Patent No. 5977336
: GENERAL INFORMATION:
: APPLICANT: Barenkamp, Stephen J
: TITLE OF INVENTION: High Molecular Weight Surface Proteins
: TITLE OF INVENTION: Of No. 5977336-Typeable Haemophilus
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Maltare, Ltd.
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/617,697
: FILING DATE: 01-APR-1996
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/302,832
: FILING DATE: 05-OCT-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US PCT/US93/02166
: FILING DATE: 16-MAR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Berkstresser, Jerry W
: REGISTRATION NUMBER: 22,651
: REFERENCE/DOCKET NUMBER: 1038-557
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0813
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1600 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-617-697-10

alignment_scores:
: Quality: 116.50 Length: 343
: Ratio: 0.613 Gaps: 19
: Percent Similarity: 55.394 Percent Identity: 21.574

alignment_block:
US-09-839-894-9 x US-08-617-697-10 ..
Align seq 1/1 to: US-08-617-697-10 from: 1 to: 1600

142 CATATATATTAATAATCAATATTAACAGATYACAGTGAAGATCACTCT 191
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
816 TyrAsnGluTyrSerLysHis.....AlaIleAsnSerSerHisAsnLe 830
192 GTATGATAGATGACTTTTATGTTGCTTCTCTCAACAATACACTTAAG 241
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
830 u.....ThrIleLeuGlyGlyAsnValThrLeuGlyG 841
242 GACCATGTCACACAGAGAGATCAATCTAGCAGTCA..... 276
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
841 Ty.....GluAsnSerSerSerSerIleThrGlyAsnIle 852
277 TCGGTGACGCGTGAACAATATTAACATTACATTT..... 312
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
853 AsnIleThrAsnLysAlaAsnValThrLeuGlnAlaAspThrSerAsnSe 869

```

```

313 .....ACGMAAAAGAGCTTTAATAAAGAGAGCTACAAATTA 352
      |||||
869 rsnthrclyleuylslysgthleuthrleucllysniileservalg 886
      |||||
333 AAGCTTAACAAATTAATGTTCAAAAGTGTAACTGCCATCCGGCTTA 402
      |||||
886 lugalysleuSERleuthrclglyalaasnaalevalglyasnleu 902
      |||||
403 ACACCT.....AAGCTGCTCATTTAACTGTA 431
      |||||
903 Serilealagluaspserthrphelysglyualaseraspalsleuvs 919
      |||||
432 T.....AAAACGGCGCTTCAGTGCAGATTATAT 463
      |||||
919 nilethrclyleuthrphethrasnslgthralasnileasnileysg 936
      |||||
464 TATATATTCCTGCTGCGAGCTAAATAATTCCTTTGGTGTATCTGG 513
      |||||
936 lvalvalysleucllysplleasnsls.....glygly..... 948
      |||||
514 GATGCTACTGAGTTAAGAGTAAAGACATATAGTACAGCTATAG 563
      |||||
949 .....leuasnilethrthrasnlasersgltglnlly 960
      |||||
564 AACT...TACATAAATATACATATTAATTAACGATAGGAATA 610
      |||||
960 sthrilleasnslgthrasnilethr.....asnsluysglsyapL 974
      |||||
611 TTCAGATATGTTTACCTGCTCAAAAGTGAAGCTGCGGTGATCTT... 657
      |||||
974 euasnile.....lysasnileysalaaspalagluliegllely 988
      |||||
658 ...AAGTTCCTCCACTGCTGGGGCACATATATTGGAAGAAATCTGT 704
      |||||
989 glyasnileserglnlysgllyasnleuthrileseraserlysva 1005
      |||||
705 TGATATGCTTTTATGATGATATAGTACAGCAGAGCTCTTGGAGA 754
      |||||
1005 lasnile.....thrasnilethrilleysa 1015
      |||||
755 TAAGATTTTACG.....GATAACAATCTTAATCTGATGGAGAA 792
      |||||
1015 laglyvaliglycllyargserserserglualagluasnlasn 1031
      |||||
793 TTTTATCTAAGAGAAATATGATGACACCAAGAAATGCAATATCTT 842
      |||||
1032 leuthrilleglin.....thrlsgluleuylsleuvalagl 1043
      |||||
843 GTCACCTCTCTGGCGGT.....AAAGTTTAACTCCACAATG 883
      |||||
1043 yaspneusnileserglypheasnlyalagluilethrlealysasn 1060
      |||||
884 GAAGCTCAATTAATATGCTGACGAGCTCTCTGGAACAACATGGAAT 933
      |||||
1060 lyseraspleuthrillegllyasnlasersgltglsalaspalays 1076
      |||||
934 AGAATTACAGCTGCACCATGCCAGAAATCAAGTGTCCGGTGTGTG 983
      |||||
1077 lys.....valthrphesplysvalysaspserslleserth 1090
      |||||
984 GCGTGA...GGTTGCAATGGATGCAAAAGTGGAATCCCGGCGGTG 1030
      |||||
1090 rAspClYhIsAnValThrLeuAsnSerGluValYsThrSerAsnGls 1107
      |||||
1031 GACATATATGCTAATATTAATGTTACT 1059
      |||||
1107 eraserAsnAlagLYsAsnAspAsnserthr 1116

```

seq.name: /cgn2.6/ptodata/2/laa/6A_COMB.pep:US-08-362-525-2

seq_documentation_block:
; Sequence 2, Application US/08362525

```

Patent No. 6027910
GENERAL INFORMATION:
APPLICANT: KITS, FRANCISCUS M.
APPLICANT: SCHREUDER, MARTEN P.
APPLICANT: TOSCHKA, HOLSER Y.
APPLICANT: VERRIPS, CORNELIS T.
TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ENZYMES TO THE
TITLE OF INVENTION: CELL WALL OF A MICROBIAL CELL BY PRODUCING A FUSION
NUMBER OF INVENTION: PROTEIN
CORRESPONDENCE ADDRESSES:
ADDRESS: CUSHMAN DABRY & CUSHMAN, L.L.P.
STREET: 1100 New York Avenue, N.W.
City: Washington
STATE: D.C.
COUNTRY: U.S.A.
Zip: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/362,525
FILING DATE: 04-JAN-1995
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: EP 92202080.5
FILING DATE: 08-JUL-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: EP 92203899.7
FILING DATE: 14-DEC-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/EP93/01763
FILING DATE: 07-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N.
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 213289/T7020(V)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 861-3000
TELEFAX: (202) 822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEO ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 650 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-362-525-2

```

```

alignment_scores:
Quality: 114.00 Length: 469
Ratio: 0.573 Gaps: 23
Percent Similarity: 42.431 Percent Identity: 20.896

```

alignment_block:

US-09-839-894-9 x US-08-362-525-2 ..

```

Align seg 1/1 to: US-08-362-525-2 from: 1 to: 650
10 ATTATTTATTTTACATGTTTCTCTGAGTACTT..... 48
|||||
8 lleuthrleupheserleuAlaLeuAlaSerAlaIleAsnIleAsn 24
|||||
49 TTTATTTGCT.....GTATCGGCGCATTAATTTT 79
|||||
24 pilerthpheserasnleuclgulllethrProleuthrAlaasnlysglnp 41
|||||
80 CCGGA.....GATGAAGCATTAATTAATTTT 108
|||||
41 roAspGlnGlyThrphAlaThrphAspPheSerIleAlaAspAlaSer 57

```

```

772 AAT.....CCTAATCTGATGGCAAAATTTTAACTAAGCAAAATTAATGCA 815
||| |||
326 AsnLeuGlyThrIaSerAlaIysSerSerPheIleSerThrThrThr 342
816 TGACACCAAGAAAT.....GCATTAAGTTTGCA..... 846
||| ||| ||| |||
342 AspLeuThrSerIleAsnThrSerAlaIysSerThrGlySerIleSerT 359
847 .....CTTCTC 852
359 hrValGluThrGlyAsnArgThrThrSerGluValIleSerHisVal 375
853 TTGGCGGTAAAGTTTAACTCCACCAAAATGACAGCTAAATATTTGC 902
||||| :||| ||| ||| ||| |||
376 ThrThrSerThrIysLeuSerProThrAlaThrThrSerLeuThrIleAl 392
903 TGACGCAGCTCTCTGGAACAACTGAAT..... 933
||||| :||| |||
392 aglInThrSerIleYrSerThrAspSerAsnIleThrValGlyThrAspI 409
934 .....ACAAT 939
409 leHISThrThrSerGluValIleSerAspValGluThrIleSerArgGlu 425
940 ACAGCTGTACACCATGCCAGAAATCAGTTCGCGGTGTGTGGCGCTG 989
||||| ||| :||| ||| |||
426 ThrIaSerThrValAlaIaIaProThrSerThrThrGlyrphrG 442
990 ACGTTTG 996
||| :|||
442 yAlaMet 444

```

seq_documentation_block:

Patent No. 5928651

; GENERAL INFORMATION:

GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J

APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins

TITLE OF INVENTION: High Molecular Weight Surface Protein of No. 5928651-Tyresable Haemophilus

TITLE OF INVENTION:	of No. 5928651-Typeable Haemophilus
NUMBER OF SENTENCES:	10

```

; NUMBER OF SEQUENCES: 10 |

```

CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Mattare, Ltd.

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

STREET: 2001 JEFFERSON HWY., 1203 CRYSTAL PLAZA
BLDG. 1

STREET: Bldg. 1
CITY: Arlington

CITY: Arlington
STATE: Virginia

STATE: Virginia
COUNTRY: U S A

```
COUNTRY: U.S.A.
```

ZIP: 22202-0286

COMPUTER READABLE FORM: |

MEDIUM TYPE: Floppy disk

```

;
;
; MEDIUM TYPE: 5 1/4" disk
COMPUTER: IBM PC compatible
;
;

```

```

;
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1 0 Version #1 30
;

```

```
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT ADDITION DATA:
```

```

; CURRENT APPLICATION DATA: 1

```

APPLICATION NUMBER: US/08/728,470

FILING DATE: 02/08/2017

CLASSIFICATION: 424

PRIOR APPLICATION DATA: 424

; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: IIS 08/302-832

APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-MAR-1993

FILING DATE: 16-MAR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US PCT/US93/02166

FILING DATE: 16-MAR-1993

;; FILING DATE: 10 MAR 1993
PRIOR APPLICATION DATA: |

APPLICATION NUMBER: GB 9205704.1

APPLICATION NUMBER: GB 9205/04.1
FILING DATE: 16-MAR-1992

FILING DATE: 16-MAR-1992

ATTORNEY/AGENT INFORMATION:

NAME: Berkstresser, Jerry W

REGISTRATION NUMBER: 22,651

REFERENCE/DOCKET NUMBER: 1038-633

REFERENCE/DOC# NO. 1030 033

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ. ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1529 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-728-470-10

alignment_scores:

Quality: 111.50 Length: 348
Ratio: 0.596 Gaps: 18
Percent Similarity: 53.736 Percent Identity: 20.690

alignment_block:

US-09-839-894-9 x US-08-728-470-10 ..

Align seg 1/1 to: US-08-728-470-10 from: 1 to: 1529

142 CATATATATTAATAACCATATTACAGCATGAGTAACTACTCT 191
:::|||||:::|||||:::|||||:::|||||:::|||||
744 TyrAsnGluYrSerLysHis.....AlaIleAsnSerSerHisAsnLe 758
192 GTATGATAGATGACTTTTATGTTGTCTTCACATACACTTATG 241
|:::|||||:::|||||:::|||||:::|||||:::|||||
758 u.....ThrIleGluGlyAsnValThrLeuGlyG 769
242 GAGCATGTCCAAACGAGTGAATCTTACAGTCA..... 276
|||:::|||||:::|||||:::|||||:::|||||
769 Ly.....GluAsnSerSerSerSerLethrGlyAsnLe 780
277 TCGGTACGGGTGAACCAATATACATTTCATTT..... 312
:::|||||:::|||||:::|||||:::|||||:::|||||
781 AsnIleThrAsnLysAlaAsnValThrLeuGlnIleAspThrSerAsnSe 797
313ACGAAAAAGAGAGTTTAAATAAAGAGAGCTACAAATTA 352
797 rAsnThrGlyLeuLysLysArgThrLeuThrLeuGlyAsnIleSerValG 814
353 AAGCCTATAAACAATTTATTTGTTCAAAAGTGTAACTGCCATCCGCGCTA 402
:::|||||:::|||||:::|||||:::|||||:::|||||
814 LugiAsnLeuSerLeuThrGlyAlaAsnIleValGlyAsnLeu 830
403 ACACCTAACCTACGCTCATTTTAACTGTAATATAAAGCGGCTTCAAGTGC 452
:::|||||:::|||||:::|||||:::|||||:::|||||
831 SerIle.....AlaGluAspSerThrPheLysGlyIleAlaSer.... 843
453 AAGTTTATTTATATATATCTCTGCTGCGAACAATAAATTTGCTTTTG 502
844AspAsnLeuAsnIleT 849
503 GTGGTATCTGG.....GATGCTACTCTGAAGTTAAGAGTAAAGACA 546
|||:::|||||:::|||||:::|||||:::|||||
849 hrGlyThrPheThrAsnAsnGlyThrAlaAsnIleAsnIleLysGlnGly 865
547 TATAGTGAGACTTATGGAACCTTACACT.....ATAATATATCAG 584
:::|||||:::|||||:::|||||:::|||||:::|||||
866 ValValLysLeuGlnGlyAspIleAsnAsnLysGlyLeuAsnIleTh 882
585 TATTAATTTAACT..... 597
882 rThrAsnAlaSerGlyThrGlnLysThrIleIleAsnGlyAsnIleThra 899
598 ..GATAAGGGAATATTCAGATATAGTTTACTCTGATCAATCAAGTACGCT 645
:::|||||:::|||||:::|||||:::|||||:::|||||
899 snGluLysGlyAspLeuAsnLe.....LysAsnIleLysAlaAspAla 913
646 GCGCTCGACTT.....ACCTGCGTCCAACTGCTGGGGGACATATAT 689
:::|||||:::|||||:::|||||:::|||||:::|||||
914 GluIleGlnIleGlyAsnIleSerGlnLysGlnGlyAsnLeuThrI 930

690 TGAAGAATAATTCGTGATATGCTTTTATGATGATGATAGTACTACAA 739
|:::|||||:::|||||:::|||||:::|||||:::|||||
930 eSerSerAspLysValAsnLe.....Thrsng 940
740 GCAGCTCTTTGGAGATTAAGATTTCAG.....GATAACAATCTT 777
:::|||||:::|||||:::|||||:::|||||:::|||||
940 InIleThrIleLysAlaGlyValGlnGlyArgSerAspSerSerGlu 956
778 AATCTGATGGGAATTTTATCTAAGGAAATTAATATGATGACACCAAGA 827
:::|||||:::|||||:::|||||:::|||||:::|||||
957 AlaGluAsnAlaAsnLeuThrIleGln.....ThrLysG 968
828 AATTCATATCTTTGTCACCTTCTTGGCGGCT.....AAAGTT 868
|:::|||||:::|||||:::|||||:::|||||:::|||||
968 uLeuLysLeuAlaGlyAspLeuAsnIleSerGlyPheAsnLysAlaGlu 985
869 TTAOTCCAAACAATGGAAGCTCATTAATATTTGCTGACGACGCTTCTGTG 918
:::|||||:::|||||:::|||||:::|||||:::|||||
985 IethrAlaLysAsnGlySerAspLeuThrIleGlyAsnAlaSerGlyGly 1001
919 GAACCAAACTGGAATAGATTTACAGCTGTCCACATGCCAGAAACAGTGT 968
:::|||||:::|||||:::|||||:::|||||:::|||||
1002 AsnAlaAspAlaLysLys.....ValThrPheAspLysValLysAs 1015
969 TCCGCTGTGTGTTGGCCTGGA...CGTTGCAATTTGATGCAAAAGTGG 1015
:::|||||:::|||||:::|||||:::|||||:::|||||
1015 pSerLysIleSerThrAspGlyHisAsnValThrLeuAsnSerGluVal 1032
1016 AAAATCCCGAGCTGACAAATATATGGTAAATATTTATGTTACT 1059
:::|||||:::|||||:::|||||:::|||||:::|||||
1032 yStrSerAsnGlySerSerAsnAlaGlyAsnAspAsnSerThr 1046
seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-719-641-10
seq_documentation_block:
; Sequence 10, Application US/08719641
; Patent No. 6218141
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/719,641
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstesser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-625
; TELECOMMUNICATION INFORMATION:

914 GluIleGlnIleGlyAsnIleSerGlnLysGluGlyAsnLeuThr1 930

TOPOLOGY: linear

MOLECULE TYPE: protein
 ORIGINAL SOURCE:
 ORGANISM: Streptococcus faecalis
 IMMEDIATE SOURCE:
 CLONE: Flg. 5a (S. faecalis)
 US-08-737-716-13

alignment_scores:

Quality:	106.00	Length:	309
Ratio:	0.716	Gaps:	17
Percent Similarity:	47.896	Percent Identity:	23.625

alignment_block:

US-09-839-894-9 x US-08-737-716-13 ..

Align seg 1/1 to: US-08-737-716-13 from: 1 to: 671

```

121 AGAAGCAATCTTCCCAACAAATATATA...TTAATAACCATATTAC 167
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
309 ArgYrAlaThrAspProSerTYrAsnAlaLysLeuAsnAsnValIleTh 325
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
168 AGCATACAGTGAAGTACATCTGTATGATAGAGTACCTTTTATGTT 217
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
325 rAlaTy.....AsnLeuThrGlnTYrAspThrProSerSerGlyAla 340
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
218 TGCTCTGCATACATTAATGAGACATGTCACACCATGAGATCCT 267
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
340 snThrGlyGlyGlyThrValAsnProGlyThrGlyGlySerAsnGln 356
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
268 AGCATGTCATCGTC.....ACGGTGAACA.....AA 296
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
357 SerGlyThrAsnThrTYrThrValLysSerGlyAspThrLeuAsnLy 373
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
297 TATACATTCATATT.....ACGAAAAAAGATTTA.... 330
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
373 sIleAlaIleGlnTYrGlyAlaSerValAlaAsnLeuArgSerTrpAsn 390
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
331 ..ATAAAAAGAGCTACAATTAAAGCTATAAACAATTATTGTCAA 378
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
390 LysLeuSerGlyAspLeuIlePheValGlyLysLeuIleValLysLys 406
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
379 AGTGTAACTGCCCATCCGCTTAACCTTAACCTCACTATTACTG 428
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
407 GlyAla.....SerGlyAsnThrGlyGlySerLysnGlyGlySe 420
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
429 TAATAAAACGCGCTCAGCTGACGACTTATATATTATATTCCTG 478
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
420 rAsnAsnAsn...GlnSerGlyThrAsnThrTYrThrValLysSer 436
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
479 GCGAAGCTAAAAATTGCTTTGGTGTATCTGGGATGCTACTCTGA 528
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
436 LysAspThrLeuAsn..... 440
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
529 TTAAGCTAAAAAGCATATAGTAGACCTATAGGACCTTACATATA 578
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
441 .....LysIleAlaIleGlnTYrGly..... 447
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
579 TATCACTATTAAATTAACTGATAAGGAAATATTCAGATATGTTAC 628
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
448 ValThrVal.....AlaAsnLeuArgSerTrp..... 456
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
629 AGTTCAAAAGTACGCTCGCTGATCTTAACTTG..... 663
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
457 .....AsnGlyIleSerGlyAspLeuIlePheValGlyLysLeu 470
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
664 .....CGTCAACTGTTGGGCGACATATATTTGAGAAATCTGT 704
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
471 IleValLysLysGlyThrSerGlyAsnThr..... 480
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
705 TGATATGTGCTTTTATGATGATATAGTACTTAACAGCAGCTCTTGG 754
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
481 .....GlyGlySerSerAsnGlyGlySer..... 488
  
```

```

755 TAAGATTTCAGATACATCTTAATCTGATGGAAATTTATCTAAG 804
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
489 .....AsnAsnAsnGlnSerGlyThrAsnThrTYrThrIle 501
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
805 AAATAATATGATGACACCAAGAAATTTGCA.....TATACTTTGTCA.. 846
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
502 LysSerGlyAspThrLeuAsnLysIleAlaIleGlnTYrGlyAlaSer 518
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
847 .....CTTCTCTGGCGG 859
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
518 lAlaAsnLeuArgSerTrpAsnGlyIleSerGlyAspLeuIlePheAla 535
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
860 GTAAGTTTAATCACTCCAAACAAATGAGCTCATTAATATTTGCGAC 909
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
535 LysGlnLysIleIleValLysGlyThrSerGlyAsnThrGlyGlySer 551
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
910 GCTTCTCTGGAACCAACTGGAATAGA 936
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
552 SerAsnGlyGlySerAsnAsnAsnGln 560
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
seq_name: /cgn2_6/plodata/2/1aa/5B.COMB.pep:US-08-728-470-9
seq_documentation_block:
  Sequence 9, Application US/08728470
  Patent No. 5928651
  GENERAL INFORMATION:
  APPLICANT: Barenkamp, Stephen J
  TITLE OF INVENTION: High Molecular Weight Surface Proteins
  TITLE OR INVENTION: of No. 5928651-Typeable Haemophilus
  NUMBER OF SEQUENCES: 10
  CORRESPONDENCE ADDRESS:
  ADDRESSEE: Shoemaker and Maltare, Ltd.
  STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
  STREET: Bldg. 1
  CITY: Arlington
  STATE: Virginia
  COUNTRY: U.S.A.
  ZIP: 22202-0286
  COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: Patentin Release #1.0, Version #1.30
  CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/728,470
  FILING DATE:
  CLASSIFICATION: 424
  PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 08/302,832
  FILING DATE: 16-MAR-1993
  PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US PCT/US93/02166
  FILING DATE: 16-MAR-1993
  APPLICATION NUMBER: GB 9205704.1
  FILING DATE: 16-MAR-1992
  ATTORNEY/AGENT INFORMATION:
  NAME: BerksStresser, Jerry W
  REGISTRATION NUMBER: 22,651
  REFERENCE/DOCKET NUMBER: 1038-633
  TELECOMMUNICATION INFORMATION:
  TELEPHONE: (703) 415-0810
  TELEFAX: (703) 415-0813
  INFORMATION FOR SEQ ID NO: 9:
  SEQUENCE CHARACTERISTICS:
  LENGTH: 1338 amino acids
  TYPE: amino acid
  STRANDEDNESS: single
  TOPOLOGY: linear
  US-08-728-470-9
  
```

alignment_scores:

Quality:	101.00	Length:	418
Ratio:	0.529	Gaps:	18
Percent Similarity:	45.694	Percent Identity:	21.531

```
alignment_block:
```

US-09-839-894-9 x US-08-728-470-9 .

Align seg 1/1 to: US-08-728-470-9 from: 1 to: 1338

```

61  GATTCGGCAGATAAAATTCGCGAGATGAAGCACTAATATTTGG 110
817  ValThrPheAspIysVal..... 822
111  CCCGCGTGCAGAGCAATCTTCCCCAAACATAATA..TTAATA 157
823  ...LysAspSerLysIleSerThrAspLysIleValThrLeuAsn 838
158  ACCAATTT..ACAGATCACTGGAAGCAATCCTATGATGAGATG 204
838  erGluValLysThrSerAsnGlySerAsnAlaGlyAsnAspAsnSer 854
205  ACTTTTATTTATGTTGTCT..... 222
855  ThrGlyLeuThrIleSerAlaIysAspValThrValAsnAsnValTh 871
223  .TCTCAATACACTTAAT.....GAGCATGCTCCACCA 256
871  rSerHisLysThrIleAsnIleSerAlaAlaIleGlyAsnValThrTr 888
257  GTGAGATCCTACCACTTCATGGTCAGCGGGAACCAATATACATTA 306
888  ysuGluGlyThrThrIleAsnAlaThrThrGlySerValGluValThrLa 904
307  CAA.....TTTAC 314
905  GluAsnGlyThrIleLysGlyAsnIleThrSerGluAsnValThrValTh 921
315  GGAAGAAAGAGATTATATATAAAGAGACTCAAAAT.....A 352
921  ratThrGluAsnLeuValThrThrGluAsnAlaValIleAsnAlaThrS 938
353  AAGGCATATAACAATATATGTTCAAAAGTGTAACGTGCCACCGGCTA 402
938  erGlyThrValAsnIleSerThrLysThrGlyAspIleLysGlyLysIle 954
403  ACACCTAACCTACGACTTATTAACGTATAAAGCGCGGTACGAGTGC 452
955  GluSerThrSerGlyAsnValAsnIle.....ThrAlaSerGlyAs 968
453  AAGTTTA..... 459
968  nThrLeuLysValSerAsnIleThrGlyGluAspValThrValThrAla 985
460  ....TATTATATATCTCGTGGCGGCACTAAAAATTTGCCCTTT 501
985  spaIaGlyAlaLeuThrThrThrAlaGlySerThrIleSerAlaThrThr 1001
502  GGGGATATCTGGGATCTCACTGGAAGTTAAGAGTAAAGACGATATAG 551
1002  GlyAsnAlaAsnIleThrThrLysThrGlyAspIleAsnGlyLysValGl 1018
552  TGAGACCTATGGAACTTACATCAATAATATCACTTAAATTAACGTGTA 601
1018  userSerSerGlySerValThrLeuValAlaThrGlyAlaThrIleuAla 1035
602  AGGGAATATTAGATATGTTACCTCACTTCAAAAGTGACGCTCGCTC 651
1035  aIcGlyAsnIleSerGlyAsnThrValThrIleThrAlaAspSerGly... 1050
652  GATCTTAACCTTCGTCACCACTGCTGGGCGCAATATATGGAAGAAATC 701
1051  ....LysLeuThrSerThrValGlySerThrIleAsnGlyThrAsnIle 1065

```

```

702 TGTGTAATGTCGGTATTATCATGAGATATACACTAACAGCAGCTTTGG 751
      |||||                               |||||||
1065 rAlITrThr.....SerSeriSeISeIISeIISeIIeG 1076
      |||||                               |||||||

752 AGATAAGATTTCAGCATTAACAAT.....CCTAATCTGATGGCAAA 792
      || ::::~::~|||::~::|||  |||
1076 lnclythrIleSerGIyAsnThrValasnValThrlalaserThrgIyasp 1092Zd
      ::::~::~|||::~::|||  |||
793 TTTTATCTA...AGGAATAATMAATGATGACACCAGAATAATTCATATAC 839
      ::::~::~|||::~::|||  |||
1093 LeuthrIleolYasnsereIatLyvalGluaLlalyasngclYalaIatrh 1109
      ::::~::~|||::~::|||  |||
840 TTTTGCA.....CTTCCTTGCGCGGTAAAGTTAA 871
      |||||::~::~|||::~::|||  |||
1109 rLeuthrIlaglUsenGclYlysLeuthrTrhgInThgIyserIerIet 1126B
      ::::~::~|||::~::|||  |||
872 CTCACCAAAAATGAGAACGCTATTAAATATTCCTGCAGCAGCTTCCTGCAA 921
      || :|||::~::~|||  |||
1126 hsrserIasnGclYInThrThrLeuthrIlalyaspsereIlela 1142Zd
      || :|||::~::~|||  |||
922 ACAACTGGAAT....AGAATTCAGCGTGCACCATGCCAGAAATCAG 965
      || ||||  |||  |||
1143 GlYAsnlleAsnaIaaIaasnValThrLeuasnthrrhgIyThrLeutH 1159
      |||  |||  |||  |||
966 TGGT.....CCGCTGTGTGTGTGTCGCCGACGTTGCCAATGG. 1002Zd
      .....  |||  |||  |||
1159 rThrThrgIyAspserrLyIsleAsnIathnrsercIyThrLeuthrIlea 1176B
      .....  |||  |||  |||
1003 .....GATGCAAAAGTGGAATAATCCAGAGCTGCA..... 1032Zd
      |||||  |||||  |||||  |||||
1176 snalalyaspaIatLyLeuasGclYlalaIalasercIyAsparGThrVal 1192Zd
      .....  |||||  |||||  |||||
1033 .....CAATATANGGTAAATATTAATGTTACTTCTTCACACCAAG 1070B
      .....  |||||  |||||  |||||
1193 ValasnIatThrsnalaserGIysenGclYAsnValThrlalalystrHse 1209B
      .....  |||||  |||||  |||||
1071 TAGT 1074
      ||||
1209 rSer 1210
      ||||

seq_name: /cgn2_6/p/tdata/2/iaa/6B-COMB.pep:US-08-719-641-9
seq_documentation_block:
Sequence 9, Application US/08719641
Patent No. 6218141
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteinsms
NUMBER OF SEQUENCES: Of No. 6218141-Typeable Haemophilus
CORRESPONDENCE ADDRESS:
ADDRESSSEE: Shomeker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
          Street: Bldg. 1
          City: Arlington
          STATE: Virginia
          COUNTRY: U.S.A.
          ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/719,641
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166

```

FILING DATE: 16-MAR-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: GB 9205704.1
 FILING DATE: 16-MAR-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Berkstresser, Jerry W
 REGISTRATION NUMBER: 22,651
 REFERENCE NUMBER: 1038-625
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 415-0810
 TELEFAX: (703) 415-0813
 INFORMATION FOR SEQ ID NO: 9:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1338 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-719-641-9

alignment_scores:
 Quality: 101.00 Length: 418
 Ratio: 0.529 Gaps: 18
 Percent Similarity: 45.694 Percent Identity: 21.531

alignment_block:

US-09-839-894-9 x US-08-719-641-9 ..

Align seg 1/1 to: US-08-719-641-9 from: 1 to: 1338

```

61  GPTACGGCAGATAAATTCGCCGAGATGAAGAAGCATATATATTTTGG 110
    |||:|:| |||:|:|:|
817  ValThrheaspIysVal..... 822
111  CCCGCGTAGAGNAGCATCTTCCCAACATATATA...TTAATA 157
    :|:|:| :|:|:|:|:|:| |||:|:|:| |||:|:|
823  ...LysAspSerLysIleSerThrAspGlyHisAsnValThrLeuAsn 838
158  ACCATATT...ACAGCATACAGTGAAGTCATCTGTATGATAGATG 204
    :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
838  erGIuValIysThrSerAsnGlySerSerAsnAlaGlyAsnAspAsnSer 854
205  ACTTTTATGTGTGCT..... 222
    ||| ||| |||
855  ThrGlyLeuThrIleSerAlaLysAspValThrValAsnAsnAsnValThr 871
223  .TTCACAAATACACTTAAT.....GGAGCATGTCCACCA 256
    |||:|:|:|:|:|:|:|:|:| ||| |||
871  rSerHisLysThrIleAsnIleSerAlaIleAlaGlyAsnValThrThrL 888
257  GTGGAATCCTAGCAGTTTCATGGTCAAGCGTGAACAAATATTAACATTA 306
    :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
888  ysgIuGlyThrThrIleAsnAlaThrThrGlySerValGIuValThrAla 904
307  CAA.....TTTAC 314
    ||| |||
905  GlInsngIyThrIleLysGlyAsnIleThrSerGIAsnValThrValThr 921
315  GGAAGAAAGAGTTTAATAAAGAGACTACAAAT.....A 352
    | :|:|:|:|:|:|:|:|:| ||| :|:|:|
921  rAlaThrGIAsnLeuValThrThrGIAsnAlaValIleAsnAlaThrS 938
353  AAGGCTATTAACATTAATGTTCAAAAGTGAATCCATCCGCGGCTA 402
    :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
938  erGIyThrValAsnIleSerThrLysThrGIAspIleLysGIyGlyIle 954
403  ACACTTAACAGCTCATTTTAATGTAATAAAGCCGCTTCAAGTCG 452
    :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
955  GluSerThrSerGIAsnValAsnIle.....ThrAlaSerGIAsn 968
453  AAGTTTA..... 459
    :|:|:| |||
968  nThrLeuLysValSerAsnIleThrGIAsnAspValThrValThrAla 985

```

```

460  ....TATTATATTCCTGCGGAGACTAAAAATTCCTTT 501
    ||| |||:|:|:|
985  spAlaGlyAlaLeuThrThrThrAlaGlySerThrIleSerAlaThrThr 1001
502  GGTGTATCTGGATGCGACTCGACTGAGTTAAGCTAAAGACGATATAG 551
    |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1002  GIAsnAlaAsnIleThrThrLysThrGlyAspIleAsnGlyLysValGI 1018
552  TGAGACCTATGCACTTACATATATATCATTATTAATTAATCAGTA 601
    :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1018  uSerSerSerGIySerValThrLeuValAlaThrGlyAlaThrLeuAlaV 1035
602  AGGCAATATTTCAGATATGTTACCTCAGTTCAAAAGTGACCTCGCTC 651
    |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1035  alGIyAsnIleSerGIyAsnThrValThrIleThrAlaAspSerGIy... 1050
652  GATCTTAACCTTGCGTCCAGCTGCGGGGACATATATTTGAAGAATTC 701
    :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1051  ....LysLeuThrSerThrValGIySerThrIleAsnGlyThrAsnSe 1065
702  TGTGATATGTGCTTTATGATGATATAGTACTAACAGACGCTCTTGG 751
    ||| |||:|:|:|:|:|:|:|:|:|:|:|:|:|
1065  rValThrThr.....SerSerGIAsnSerGIyAspIleG 1076
752  AGATAAGATTTCAGCATACAT.....CCTAATCTGATGGCAA 792
    || :|:|:|:|:|:|:|:|:| ||| |||
1076  IuGIyThrIleSerGIyAsnThrValAsnValThrAlaSerThrGIyAsp 1092
793  TTTATATCA...AGAAATTAATGATGACACCAAGAAATTCATATAC 839
    :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1093  LeuThrIleGIyAsnSerAlaLysValGIuAlaLysAsnGlyAlaAlaTh 1109
840  TTTGTCA.....CTTCCTTGCGGCGTAAAGTTTAA 871
    |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1109  rLeuThrAlaGIuSerGIyLysLeuThrThrThrGIySerSerIleT 1126
872  CTCACAAATATGACGTCATTAATATGCTGACGCCAGCTCTCTGGA 921
    || :|:|:|:|:|:|:|:|:| ||| :|:|:|:|
1126  hrSerSerAsnGIyGIuThrThrLeuThrAlaLysAspSerSerIleAla 1142
922  ACAACGTGAAT.....AGATATACAGCTGACCATGCGACGAATACAG 965
    ||| ||| |||
1143  GIyAsnIleAsnAlaAlaAsnValThrLeuAsnThrThrGIyThrLeuTh 1159
966  TGTG.....CCGCTGTGTGTGCTGCGACGTTTGCATATG 1002
    :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1159  rThrThrGIyAspSerLysIleAsnAlaThrSerGIyThrLeuThrIleA 1176
1003  ....GATGCAAAAGTGAAATCCCGAGCTGCA..... 1032
    |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1176  snAlaLysAspAlaLysLeuAspGIyAlaAlaLysSerGIyAspArgThrVal 1192
1033  ....CAATATATGGTAAATATTAATGTTACTTTCACACCAAG 1070
    :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1193  ValAsnAlaThrAsnAlaSerGIySerGIyAsnValThrAlaLysThrSe 1209
1071  TAGT 1074
    ||| |||
1209  rSer 1210

```

seq_name: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:US-08-617-697-9

seq_documentation_block:

; Sequence 9, Application us/08617697

; Patent No. 5977336

; GENERAL INFORMATION:

; APPLICANT: Barenkamp, Stephen J

; TITLE OF INVENTION: High Molecular Weight Surface Proteins

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Shoemaker and Mattare, Ltd.

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/617,697
FILING DATE: 01-Apr-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 05-Oct-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-557
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1599 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-617-697-9
alignment_scores:
Quality: 101.00 Length: 418
Ratio: 0.529 Gaps: 18
Percent Similarity: 45.694 Percent Identity: 21.531
alignment_block:
US-09-839-894-9 x US-08-617-697-9 ..
Align seg 1/1 to: US-08-617-697-9 from: 1 to: 1599
61 GATCGGACGATTAATTCCTCCGGAAGATGAAGCATTAATATTTTGG 110
|||||
1077 ValThrPheAspLysVal..... 1082
111 CCGCGGACGAGACGATCTCTCCCAACATATATATA..TTAATA 157
|||||
1083 ...LysAspSerLysLieserThrAspGlyHisAsnValThrLeuAsn 1098
158 ACCATAT...ACAGCATACGTAAGTCACTGTATAGTAGATG 204
|||||
1098 ergLysValLysThrSerAsnGlySerSerAsnLagLysAsnAspAsnSer 1114
205 ACCTTTTATGTTGCT..... 222
|||||
1115 ThrGlyLeuThrLieserAlaLysAspValThrValAsnAsnValThr 1131
223 ..TCGCAATACACTTAAT.....GGAGCATGTCACCA 256
|||||
1131 rSerHisLysThrLieserAlaLysAsnValThrThrL 1148
257 GTGAGATCTTACGATCGTCAGCGGTGAACAATATATACATTA 306
|||||
1148 ysgLugLysThrThrLiesAsnAlaThrThrGlySerValGluValThrAla 1164
307 CAA.....TTTAC 314
|||

1165 GluAsnGlyThrLiesGlyAsnLieserSerGlnAsnValThrValTh 1181
315 GGAAGAAAGAGTTTATTAATAAGAGCATCAAAAT.....A 352
|||||
1181 rAlaThrGluAsnLeuValThrThrGluAsnAlaValLiesAsnAlaThs 1198
353 AAGGCTATAAACAATATGTTCCAAAGTGTAACTGCCATCCGCGCTA 402
|||||
1198 ergLysThrValAsnLieserThrLysThrGlyAspLiesGlyGlyLe 1214
403 ACACCTTAACACGCTCATTTTAACTGTAATAAAGCGGCTTCAGTGC 452
|||||
1215 GluSerThrSerGlyAsnValAsnLies.....ThrLieserGlyAs 1228
453 AGCTTA..... 459
|||||
1228 nThrLeuLysValSerAsnLieserThrGlyGlnAspValThrValThAla 1245
460TATTTATATATTCCTGCGGCACTAAATAATTTGCCCTTT 501
|||||
1245 spAlaGlyAlaLeuThrThrThrAlaGlySerThrLieserAlaThrThr 1261
502 GGTGATATCGGATGCTACTGGAAGTTAAGCTAAAAAGAGATATAG 551
|||||
1262 GlyAsnAlaAsnLieserThrThrGlyAspLiesAsnGlyLysValG 1278
552 TGAGACCTATGAGACTTACCTATTAATATACATAATTAACGATA 601
|||||
1278 userSerSerGlySerValThrLeuValAlaThrGlyAlaThrLeuAla 1295
602 AGGAAATATTCAGATATGTTACTCCTCAAGTCAAAAGTACGCTCCGCT 651
|||||
1295 alGlyAsnLieserGlyAsnThrValThrLieserAlaAspSerGly... 1310
652 GATCTTAACCTGCGTCCCACTGGTGGGACATATATTGGAAGAAATTC 701
|||||
1311LysLeuThrSerThrValGlySerThrLiesAsnGlyThrAsn 1325
702 TGTGATATGCTGTTTATGATGATATAGTACTAACACAGCGCTTTGG 751
|||||
1325 rValThrThr.....SerSerGlnSerGlyAspLies 1336
752 AGATTAAGATTTCAGATTAACAAT.....CCTAATCTGATGGGAAA 792
|||||
1336 LucGlyThrLieserGlyAsnThrValAsnValThrLieserThrGlyAsp 1352
793 TTTTATCTA...AGAAATTAATGATGACACCAAGAAATTCATATAC 839
|||||
1353 LeuThrLiesLysAsnSerAlaLysValGluAlaLysAsnGlyLysAla 1369
840 TTTGTCA.....CTTCTCTGCGGCGTAAAGTTTAA 871
|||||
1369 rLeuThrAlaGluSerGlyLysLeuThrThrGlnThrLieserSerLies 1386
872 CTCCAACAAATGAGCGTCAATTAATGTCGACGACGCTTCGTGGA 921
|||||
1386 hrSerSerAsnGlyGlnThrThrLeuThrAlaLysAspSerSerLiesAla 1402
922 ACAACTGCAAT...AGAAATTACAGCTGCACCATGCAACAATCAG 965
|||||
1403 GlyAsnLiesAsnAlaLysAsnValThrLeuAsnThrThrGlyThrLeuTh 1419
966 TGT.....CCGCTGTTGTTGGCTGCGACGTTTCATTTG 1002
|||||
1419 rThrThrGlyAspSerLysLiesAsnAlaThrSerGlyThrLeuThrLies 1436
1003GATGCAAAAGTGAATAATCCGAGGCTGA..... 1032
|||||
1436 snLysAspAlaLysLysAspGlyAlaLieserGlyAspArgThrVal 1452
1033CATATATGCGTAAATATATATGTTACTTTCACACCAAG 1070
|||||
1453 ValAsnAlaThrAsnAlaSerGlySerGlyAsnValThrAlaLysThrSe 1469

TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein

NUMBER OF SEQUENCES: 5

RESPONSE ADDRESS:
ADDRESS: Flehr, Hohbach, Test, Albritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10661A
FILING DATE: 16-AUG-1995
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/296,791
FILING DATE: 25-AUG-1994

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Treccarlin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: FP-59941/RFT
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEO ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1848 amino acids
TYPE: amino acid
TOPOLOGY: unknown
PCT-US95-10661A-6

alignment_scores:
Quality: 99.00 Length: 325
Ratio: 0.593 Gaps: 20
Percent Similarity: 51.385 Percent Identity: 23.385

alignment_block:
US-09-839-894-9 x PCT-US95-10661A-6 ..
Align seg 1/1 to: PCT-US95-10661A-6 from: 1 to: 1848

```
55 TTTGGCTGATCGGAGATATAATTCGCGAGATGAACATTAATATAT 104
||| : : : : : : : : : : : : : : : : : : : : : : : :
753 PheLysAlaThrThrMetAsnValThrGlyAsnAlaSerLeuYrSer.. 768
105 TTTTGGCCCGCGTGACAGAAAGATCTCCCAACATTAATATATTA 154
||| : : : : : : : : : : : : : : : : : : : : : : : :
769 .. : : : : : : : : : : : : : : : : : : : : : : : :
155 ATTAACCATATACGATACAGTGAAGT.....CATACTGTGATGAT 198
: : : : : : : : : : : : : : : : : : : : : : : :
777 eirAsn...IleThrAlaSerAsnAsnAlaGlnValHisIleGlyThrLys 792
199 AGGATGACTTTTATGTGCTCTCT..... : : : : : : : : : : 225
793 ThrGlyAspThrValCysValArgSerAspYrThrGlyYrValThrCy 809
226 .. : : : : : : : : : : : : : : : : : : : : : : : :
809 sHisAsnSerAsnLeuSerGluYrAlaLeuAsnSerPheAsnPro.... 824
272 GTTCAATCGGTGACGGTGAACAATAATTAACA.....TTACAATTAG 315
: : : : : : : : : : : : : : : : : : : : : : : :
825 .. : : : : : : : : : : : : : : : : : : : : : : : :
316 GAAAAAAGAGTTTAATAAAGAGAGAGATCAAAATTAAGCTTAACA 365
841 LeuGlyLysAlaAsnLeuPheGlyThrIleGlnSerIleGlyThrSerG 857
```

```
366 ATTAATGTTCAAAAGTGTAACTGCCATCGGCGCTAACAATTAACTAC 415
||| : : : : : : : : : : : : : : : : : : : : : : : :
857 nValAsnLeuLysGlu...AsnSerHisThrPheLeuThrGlyAsnSer.. 872
416 CTCAATTTAACTGTAAATMAA...AACGGCGCTTCAAGTGTATATAT 462
||| : : : : : : : : : : : : : : : : : : : : : : : :
873 .. : : : : : : : : : : : : : : : : : : : : : : : :
463 TTATATATTCCTGCTGGCAACTAAAAATTGGCTTTGGTGTATCTG 512
: : : : : : : : : : : : : : : : : : : : : : : :
884 IleHisLeuAsnAlaGlnAsnAspAlaAsn..... : : : : : 893
513 GGATGCTACTCTGAGATTAAAGTAAAGAGATATAGTGAGACTATG 562
894 .. : : : : : : : : : : : : : : : : : : : : : : : :
563 GAACCTTCACTATATAATATCACTTAAATTAAGTGAAGGAATATTT 612
: : : : : : : : : : : : : : : : : : : : : : : :
899 snThrLeuThrValAsn.....SerLeuSerGlyAsnGlySerPhe 912
613 CAGATATGTTAGCTCACTCAAGTTCAAAAGTGACGCT...CGCGCTCTTAA 659
913 TyrTyrThrPValAspPheThrAsnAsnLysSerAsnLysValValAs 929
660 CTTCGCTCAACTGCTGGGGGCACATAT.....ATTGAA 694
929 nLysSerAlaThrGlyAsnPheThrLeuGlnValAlaAspLysThrGly 946
695 GA...AATCTGTGATATGCTTTTATGATGATGATTAAGTAAACAC 741
: : : : : : : : : : : : : : : : : : : : : : : :
946 IupProAsnHisAsnGlnLeuThrLeuPheAspAlaSerAsnAlaThrArg 962
742 AGCTCTTGGAGATAGATTTCAAGTAAACAT...CCTAAATCTGATG 788
963 .. : : : : : : : : : : : : : : : : : : : : : : : :
789 GAATTTTATCTAAGAAATTAATGATGACACCCAAAGAAATTGCATATA 838
||| : : : : : : : : : : : : : : : : : : : : : : : :
979 pLysTyrLysLeuArgAsnValAsnGly.....ArgTyrAspLeuYrA 994
839 CTTCGCTCACTTCTTGGCGGGTAAAGTTTAAGTCAACAATGAGACG 888
: : : : : : : : : : : : : : : : : : : : : : : :
994 snProGluValGlnLysArgAsnGlnThrValAspThrThrAsnIleThr 1010
889 TCATTAAT.....ATTGCTGACGACGCTTCCTGGAAACAACCTGAA 932
1011 ThrProAsnAspIleGlnAlaAspAlaProSerAlaGlnSerAsnAsnG 1027
933 TAGAATTACAGCTGTCACATGCA 957
1027 uGlnIleAlaArgValGlnThrPro 1035
```

seq_name: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:US-08-409-995-4

seq_documentation_block:
Sequence 4, Application US/08409995
Patent No. 5646259

GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen I.
APPLICANT: St. Geme III, Joseph W.
TITLE OF INVENTION: Haemophilus Adhesion Proteins
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSER: Flehr, Hohbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/409,995
FILING DATE: 24-MAR-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-61053/RFT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1912 amino acids
TYPE: amino acid
STRANDEDNESS: double
TOPOLOGY: unknown
US-08-409-995-4

```

```

alignment_scores:
Quality: 94.50 Length: 310
Ratio: 0.602 Gaps: 14
Percent Similarity: 50.645 Percent Identity: 19.677

```

alignment_block:

US-09-839-894-9 x US-08-409-995-4

Align seg 1/1 to: US-08-409-995-4 from: 1 to: 1912

```

232 ACACCTTAATGAGACATGTCACCAAGTGAGATCCTAGCAGTTCAATCGT 281
||||| |||||
805 ThrilegylgYAsnThrProthr.....GlyGlyThrThrAl 817
282 CAGCGGTAAACAAATATACATTCATTTACGAAAAAGAACTTTAA 331
||||| |||||
817 aThrProlysnValasnThr.....SerThrAlasnspGlyLeuasnPha 833
332 TAAAGAAGACCTACAAATTAAGGCTTAACATTTATTTGTTCAAAAGT 381
||||| |||||
833 lalysglThrAlasnspalaserGlySerlysnValTyrleuysglY 849
382 GTTAACGTC.....CCATCCGCGCTTAACACTTAACGACGTCGA 419
||||| |||||
850 lIeaIaThrThrleuThrGluProserAla..GlyAlalysSerSerH 865
420 TTTTACTGTAATAAAACCG.....GCTTCAGGTGCAAGTTTAT 460
||||| |||||
865 sValaspLeuasnValaspAlaThrlySerSerasnAlaIaSerIleG 882
461 ATTATATATTCCTGCT.....GGCAACTAAAAATTTG 495
||||| |||||
882 lAspValleuThrGlyAlaGlyTrrpasnIleGlnGlyasnGlyAsnVal 898
496 CTTTGTGCTATCTGGATGCTACTCTGAAGTTAAGTAAAAAGACG 545
||||| |||||
899 AspYrValAlaThrTyrAspThrVal.....AspPh 909
546 ATATAGTGAGACCTATGACATTAATATATTCACCTATTAATTAA 595
||||| |||||
909 eThraspSerSerThrGlyThrThrThrValThrValThrGlnLysAla 926
596 CTGATAGGGA..... 606
|||||
926 spGlyLysGlyAlaAspValLysIleGlyAlaLysThrSerValLleLys 942
607 .....AATATTCAGATATGTTTACTCAGTTCAAAAGTGACGCTCGGT 650
||||| |||||
943 AspHisasnGlyLysLeuPherThrGlyLysAspLeuLysAspAlaAsnAs 959

```

```

651 CGATCTTAACGTC.....CGTCCACCTGGTGGGCGACAT 685
||||| |||||
959 nGlyAlaThrValSerGluAspAspGlyLysAspThrGlyThrGlyLeu 976
686 AATTTGAGAAATTCCTGTGATATGCTTTTATGATGATAT 729
||||| |||||
976 aThrAlaLysThrValIleAspAlaValasnLysSerGlyTrrpArVal 992
729 ..... 729
993 ThrGlyGluGlyAlaThrAlaGluThrGlyAlaThrAlaValasnAlaG 1009
730 .....AGTACTACAGACGCTTTGGAGATTAAGTTTCAGATTAACA 772
||||| |||||
1009 yAsnAlaGluThrValThrSerGlyThrSerValasnPhelLysasnGly 1026
773 AT.....CCTAATCTGATGAGGAAATTTATCTAAG 804
||||| |||||
1026 snAlaThrThrAlaThrValSerLysAspAsnGlyAsnIleasnValLys 1042
805 .....AAATTAATGATGACACCAAGA 827
1043 TyrAspValasnValGlyAspGlyLeuLysIleGlyAspAspLysLys 1059
828 AATTCATATATCTTGTCACTT..CTCTGGCGGGTAAATTTAATC 874
||||| |||||
1059 eValAlaAspThrThrThrThrThrValThrGlyGlyValSerValP 1076
875 CAAACAATGAGACGTCATTAAT.....ATTGTCAGCGCAGCT 912
||||| |||||
1076 roAlaGlyAlaAsnSerValasnAsnLysLysLeuValasnAlaGlu 1092
913 TCTCTGAAACAACTGGAATGAATTACA 942
||||| |||||
1093 GlyLeuAlaThrAlaLeuasnLeuSer 1102

```

seq_name: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:US-08-685-467-4

seq_documentation_block:

Sequence 4, Application US/08685467

Patent No. 6060059

GENERAL INFORMATION:

APPLICANT: St. Gene III, Joseph W.

TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert

STREET: Four Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: United States

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/685,467

FILING DATE: 22-JUL-1996

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/409,995

FILING DATE: 24-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Silva, Robin M.

REGISTRATION NUMBER: 38,304

REFERENCE/DOCKET NUMBER: A-61053-2/RFT/RMS

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 781-1989

TELEFAX: (415) 398-3249

TELEX: 910 277299

; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1912 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: unknown
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: protein
 ; US-08-685-467-4

alignment_scores:
 Quality: 94.50 Length: 310
 Ratio: 0.602 Gaps: 14
 Percent Similarity: 50.645 Percent Identity: 19.677

alignment_block:
 US-09-839-894-9 x US-08-685-467-4 ..

Align seg 1/1 to: US-08-685-467-4 from: 1 to: 1912

```

232 ACACCTAATGAGCATGTCACACAGAGATCCTACAGTTCGCGT 281
|||:||||| |||||
805 ThrIleGlyGlyAsnThrProThr.....GlyGlyThrThrAl 817
282 CAGCGGTGAACAATATATACATTACATTGCGAAAAAGATTAA 331
|||:||||| |||||
817 aThrProLyValAsnIleThr...SerThrAlaSpGlyLeuAsnPhea 833
332 TAAAAAGAGACCTCAAAATTAAGGCTATAACATTTATTGTCAAAAGT 381
|||:||||| |||||
833 IalysGluThrAlaAspAlaSerGlySerLysAsnValTyrLeuLysGly 849
382 GTTAACTGC.....CCATCGGCGCTAACACTTAACCTGACGTCGA 419
|||:||||| |||||
850 IleAlaThrThrLeuThrGluProSerAla...GlyAlaLysSerSerHi 865
420 TTTTAACCTGATATAAAACCGG.....GCTTCAGGTGCAAGTTTAT 460
|||:||||| |||||
865 sValAspLeuAsnValAspAlaThrLysLysSerAsnAlaAlaSerIleG 882
461 ATTATATATTCCTCGCT.....GGCGAACTAAAAAATTG 495
|||:||||| |||||
882 IusAspValLeuArgAlaGlyTTrpAsnIleGlnLysAsnVal 898
496 CCTTTGCGTATCTGGAGTGTCTACTCTGAAGTTAAGATAAAAGACG 545
|||:||||| |||||
899 AspTyrValAlaThrTyrAspThrVal.....AsnPh 909
546 ATATAGTGAGACCTATGAACTTACACTATAATATCATTAAATTAA 595
|||:||||| |||||
909 eThrAspAspSerThrGlyThrThrThrValThrValThrGlnLysAla 926
596 CTGATAAGGGA..... 606
|||:||||| |||||
926 spGlyLysGlyAlaAspValLysIleGlyAlaLysThrSerValIleLys 942
607 .....AATATTCAGATATGCTTACTCTCAAGTCAAAAGTACGCTCGGT 650
|||:||||| |||||
943 AspHisAsnGlyLysLeuPheThrGlyLysAspLeuLysAspAlaAsnAs 959
651 CGATCTTAACCTG.....CGTCAACTGCTGGGGGCACAT 685
|||:||||| |||||
959 nGlyAlaThrValSerGluAspAspGlyLysAspThrGlyThrGlyLeuV 976
686 ATATTGAAGAATTCGTGATATGCTTTTATGATGATAT..... 729
|||:||||| |||||
976 alThrAlaLysThrValIleAspAlaValAsnLysSerGlyTTrpArgVal 992
729 ..... 729
993 ThrGlyGluGlyAlaThrAlaGluThrGlyAlaThrAlaValAsnAlaGlu 1009
730 .....AGTACTAACAGACGCTCTTTGGAGATAAGATTTCAGGATAACA 772
  
```

```

1009 yAsnAlaGluThrValThrSerGlyThrSerValAsnPhelysAsnGlyA 1026
|||:||||| |||||
773 AT.....CCTAATCTGATGGGAAATTTTATCTTAAGG 804
|||:||||| |||||
1026 snAlaThrThrAlaThrValSerLysAspAsnGlyAsnIleAsnValLys 1042
805 .....AAATTAATGATGACACCAAGA 827
|||:||||| |||||
1043 TyrAspValAsnValGlyAspGlyLeuLysIleGlyAspAspLysLysIl 1059
828 AATGCAATATACCTTTGCTACTT...CTCTGGCGGGAATAAGTTTAACTG 874
|||:||||| |||||
1059 eValAlaAspThrThrThrThrValThrGlyLysValSerValP 1076
875 CAACAAATGGAACGTCATTAAAT.....ATTGCTGACGCGACCT 912
|||:||||| |||||
1076 roAlaGlyAlaAsnSerValAsnAsnLysLysLeuValAsnAlaGlu 1092
913 TCTCTGAACACAACTGGAATGAAATTACA 942
|||:||||| |||||
1093 GlyLeuAlaThrAlaLeuAsnLysLeuSer 1102
  
```


This Page Blank (uspto)

This Page Blank (uspto)

134 euAlaLeuAsnSerSerIspHegInCysAsnArgLugInAlaSerGly 150
135
451 GCAAGATTATATTAATTAATTCGCTGCGGACGACATAAAAAATTTGCCTTT 500
136
151 AlaTrnLeuSerLeuTrIleProAlaGlyGluLeuAsnLysLeuProPh 167
152
501 TCGTGGTATTCGCGATGCTACCTCGAAGTTAAGGTAAAAAGCAGATATA 550
153
167 eGlyGlyValTrIpaSnAlaValLeuLysLysAsnValLysArgArgTyrA 184
154
551 GTGAGACCTATGAGAACTTACACTATAATATCACTATTAATTAAGTAT 600
155
184 sPTTrTrTyrcIyTrIyTrIyThrIleAsnIleTrnValAsnLeuTrnAsp 200
156
601 AAGGAAATATTCAGATATGCTTACCTCAGTTCAAAAGTACCGTCGCC 650
157
201 LysGlyAsnIleGlnIleTrIleProGlnIlePheLysSerAsnAlaArgVa 217
158
651 GCATCTTAACTTGGCTCCACTGGGCGGCGACATATATGGAAGAAAT 700
159
217 LAspLeuAsnLeuArgProTrnGlyGlyGlyTrIyTrIleGlyArgGlns 234
160
701 CTGTTGATATGTGCTTTATGATGATATAGTACTTAACAGACGCTCTTG 750
161
234 eValLAspMetCysPheTyrAspGlyIySerTrnAsnSerSerLeu 250
162
751 GAGATTAAGATTTCAGGATACAAATCTTAATCTGATGGGAAATTTATCT 800
163
251 GlnIleArgPheGlnAspAspAsnSerLysSerAspGlyLysPheTyrLe 267
164
801 AAGGAAATTAATGATGACACCAAGAAATGCAATACCTTGTGCACTTC 850
165
267 uLysLysIleAsnAspSerLysGluLeuValTyrThrLeuSerLeuL 284
166
851 TCTTGGCGGGTAAAGTTAACTCCACAAATGAGACGTCATTAATAT 900
167
284 euLeuAlaGlyLysAsnLeuTrnProTrnAsnGlyGlnAlaLeuAsnIle 300
168
901 GCTGACGACGCTTCTCTGGAAACAACATCGGAATTAAGATTACAGCTG 950
169
301 ...AsnTrnAlaSerLeuGlnTrnAsnTrpAsnArgIleThrAlaValTh 316
170
951 CATGCGCAAAATCAAGTTCGCGGTGTTGCTGTTGGCCGACGTTTGCAT 1000
171
316 TrnProGlnIleSerValProValLeuCysTrpProGlyArgLeuGlnL 333
172
1001 TCGATGCAAAAGTGGAAATCCGAGCGCTGGACAAATATATGGTAAAT 1050
173
333 euAspAlaLysValLysAsnProGlnAlaGlyGlnTyrMetGlyAsnIle 349
174
1051 AATGTTACTTCCACCAAGTAGTCAAAACCTC 1083
175
350 LysIleTrnPheTrnProSerSerGlnTrnLeu 360
176
seq_name: plr2:S49539
177
seq_documentation_block:
178
179 Cood protein precursor - Escherichia coli
180
181 C:Species: Escherichia coli
182
183 C:Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 10-Dec-1999
184
185 C:Accession: S49539
186
187 Mol: MicrobioL, B.J.: 387, Karakashian, A.; Melsen, L.R.; Wakefield, J.C.; Scott, J.R.
188
189 A:Title: Cood and Cood are required for assembly of CSI pilI.
190
191 A:Reference number: S49538; MUID:94344028
192
193 A:Accession: S49539
194
195 A:Status: preliminary
196
197 A:Molecule type: DNA
198
199 A:Residues: 1-363 <PRO>
200
201 A:Cross-references: EMBL:X76908; NID:g488735; PIDN:CA54230.1; PID:g488737
202
203 A:Superfamily: Escherichia coli colonizing factor antigen cfaE

Alignment_scores:		Quality: 949.50	Length: 371
Ratio: 3.230		Gaps: 11	
Percent Similarity: 79.245	Percent Identity: 53.100		
alignment_block:			
US-09-839-894-x S49539 ..			
Align seg 1/1	to: S49539	from: 1	to: 363
7	AAGATTTTATTATTATTATTACATGTGTTTCTCTGACATCTTTTACATT	56	
	:::	: :	
2	LysLysIlePheIlePheLeu.....SerIleIlePheSerIal	14	
57	TGCGTATACGGCAGATAAATTCGCCGAGATGAA.....ACATATACTA	100	
	:::	::: :	
14	aalValSerIalIaGlyArgTyrProGluThrThrValGlyAsnLeuThrL	31	
101	ATATTTT...GGCCCGCGT...GACAGGAGCATGTCTCCCAACAT	144	
::	:::	::: :	
31	ysSerPheGlnIalIaIaProArgLeuAspArgSerValGlnSerProIleTyr	47	
145	AATATATTAATTAACCATATTATACAGCATACAGTAAGTAAGTACTGTGA	134	
	:::	: :	
48	AsnIlePheThrAsnHisValAlaGlyTyrSerLeuSerHisSerLeuTyr	64	
195	TGATGAGTACACTTTTATATGTTTGTCTGTCCACAAATACACTTAATGAG	244	
	:::	: :	
64	rasPrigIleValIlePheLeuCyThrSerSerSerAsnProAlaIsmIaLys	81	
245	CATGTCCACACAGTGGAGAACTTACAGATTCATCGGTACAGC...GGTAA	291	
	:::	::: :	
81	IaCySerProThrIle.....GlyThrSerGlyValGlnTyrGlyThr	94	
292	ACAAATATATACATTACAAATTTAGCGAAAAAGAAAGTTAAATAAAAAGAA	341	
	:::	: :	
95	ThrThrIleThrIleGlnPheThrGluLysArgSerLeuIleLysArgAs	111	
342	GCTACAAATTAAGCGCTTAACAAATTAATGTTCAAAAGTGTTAAACGCC	391	
	:::	: :	
111	IleLysLeuIalIaGlyAsnLysLysProIleThrPdlAsnGlnSerGlySa	128	
392	CATCCGCG.....CTAACCTTACTACAGCTCATTTTAACTGTAAATAA	435	
::	:::	::: :	
128	sPheSerAsnLeuIaIaIaValIleuAsnSerLysSerTyrPserGlyAla	144	
436	AACGGC...GCTCAGGTGCAAGTTTATATTATATTTCTGTGTCGCGA	462	
	:::	: :	
145	HisGlyAsnAlaAsnGlyThrLeuLeuAsnLeuTyrIleProAlaGlyL	161	
483	ACTAAAAAATTTGCCTTTGGTGGAATCGGAGTACACTGTGAAGTTAA	532	
	:::	: :	
161	WileAsnLysLeuProPheGlyGlyIleThrPdlAlaIaThrLeuIleLeuA	178	
533	GAGTAAAAAGCATATATAGTAGAGCC.....TATGGAATTAC	570	
	:::	: :	
178	rgLeuSer...ArgTyrGlyGlnValSerSerThrHisTyrGlyAsnTyr	193	
571	ACTATAAATATACATATTAAATTTAACAGATATAGGAAATATTCAGATATG	620	
	:::	: :	
194	ThrValAsnIleThrThrValAspLeuThrAspLysGlyAsnIleGlnValTyr	210	
621	GTTACCTAGTTCAAAATAGCAGCTCGGGTGGATGATCTTAACCTGGGTCAA	670	
	:::	: :	
210	PleuProGlyPheHisSerAsnProArgValAspLeuAsnLeuArgProI	227	
671	CTGGTGGGGGACATATATTGGAAGAAATTCGTGTGATATGCGTTTAT	720	
	:::	: :	
227	IeGlyAsnTyrLysTyrSerGlySerAsnSerLeuAspMetCysPheTyr	243	
721	CATGATATAGTACTACAGCAGCTCTTTGGAGTAAAGATTTCAGAGTTAA	770	
	:::	: :	

gn seg 1/1 to: S49539 from 1 to: 363

```
7 AACATTATTTATTTTACATGTTTTCCTCAGACTTTTACATT 56  
   ||| :|||:||||| | | | | | | | | | | | |  
2 LysLysIlePheIleLeu.....SerIleIlePheSerIal 14  
   ::::| | | | | | | | | | | | | | | | | |  
57 TGCTGTATGGCGAAGATAATCCCGGAGANGAA.....AGCATACTA 100  
   ::::| | | | | | | | | | | | | | | | | |  
14 aValValSerAlaIglYArgTyProGIuThrThrValGlYasnLeuThrL 31  
   | | | | | | | | | | | | | | | | | | | | | |  
101 ATATTATT...GGCCGGCT+.GACAGACGAATCTCCCCCAACAT 144  
   ||| :|||:||||| | | | | | | | | | | | |  
31 ySerPheGlnAlProAlaGluAsnParYgSerValGInSerProIleTy 47  
   ||| :|||:||||| | | | | | | | | | | | |  
145 AATATATTAATTAACCATTATTACAGCATACAGTGAAAGTCATACTGTGA 194  
   ||| :|||:||||| | | | | | | | | | | | |  
48 AsnIlePheThrAshHisValAlaIglYrSerLeuSerHisSerLeuTy 64  
   ||| :|||:||||| | | | | | | | | | | | |  
195 TGAATAGAGACTTTTATTGTGTTTGCTTCACAAATACACTTAATGGAG 244  
   ||| :|||:||||| | | | | | | | | | | | |  
64 rAParGIleValAlPheLeuCySthrSerSerAsnProValAsnIylA 81  
   ||| :|||:||||| | | | | | | | | | | | |  
245 CAGTGCCAACAAGTGAGATCTTACAGCTCATCGGCAGC....GGGAA 291  
   ||| :|||:||||| | | | | | | | | | | | |  
81 lacySProHrIle.....GlyThrSerGIeValGlnTyrgLYThr 94  
   ||| :|||:||||| | | | | | | | | | | | |  
292 ACAAAATTAACATTACATTATTCGGAAAAAGAAGTTAATAAAAGAGA 341  
   ||| :|||:||||| | | | | | | | | | | | |  
95 ThrThrlleThrLeugInlPheThrclulYasrGerleuIleYsArGas 111  
   ||| :|||:||||| | | | | | | | | | | | |  
342 GCTACAAATTTAAAGGCTAATAACAATTAATGTCAAAGTGTAACTGCC 391  
   ||| :|||:||||| | | | | | | | | | | | |  
111 nIleAsnLeuAlaIglYasnLySylProIleTrpGlUasnGInserCySa 128  
   ||| :|||:||||| | | | | | | | | | | | |  
392 CATCGCGC.....CTAACACTTACAGCTCATCTTTTAAGTGAATAAA 435  
   ||| :|||:||||| | | | | | | | | | | | |  
128 spdPheSerIleuLeuEtyValIleuAsnSerLySerrTPSerCySglYala 144  
   ||| :|||:||||| | | | | | | | | | | | |  
436 AACGGC...GCTTACAGTGCACAGTTTAATTTATATATTCCTGTGGSCGA 482  
   ||| :|||:||||| | | | | | | | | | | | |  
145 HisGIYAsnAlaIsnclYthrLeuLeuAsnLeuTyrlleProAlaIglYgl 161  
   ||| :|||:||||| | | | | | | | | | | | |  
483 ACTAAAAAATTTGGCTTTGGTGTATCTGGAGATCACTCGAAGTTAA 532  
   ||| :|||:||||| | | | | | | | | | | | |  
161 uIleAsnLySleuProPheGIglYlletRpolUalathrLeuIleLeua 178  
   ||| :|||:||||| | | | | | | | | | | | |  
533 GAATAAAAAGACGATRTAGTGAGACC.....TAGGAACCTTAC 570  
   ||| :|||:||||| | | | | | | | | | | | |  
178 rgLeuSer...ArgTYrGIglYValISerSerThrHisTYrGIYAsnTytr 193  
   ||| :|||:||||| | | | | | | | | | | | |  
571 ACATATAATATACATTTAATTAATCTATTAAGGAATATTCAGATATG 620  
   ||| :|||:||||| | | | | | | | | | | | |  
194 ThrValAsnIleThrValAspLeuThrAspLySglYsnIIegIlnValTr 210  
   ||| :|||:||||| | | | | | | | | | | | |  
621 GTTACCTCAGTTCAAAAGAGAGCGCTGGCGTGATCTTAACCTGGCTCAAA 670  
   ||| :|||:||||| | | | | | | | | | | | |  
210 pleuProGIYrPheHisSerAsnProrIGrYValAspLeuAsnLeuKrpOI 227  
   ||| :|||:||||| | | | | | | | | | | | |  
671 CTGGTGGGGGCACATATATTTGGAGAAATCTGTGTATATGTGCTTTTAT 720  
   ||| :|||:||||| | | | | | | | | | | | |  
227 leGIYsnTytrLySYrSerGIYSerIleSnSerLeuAspMecySpHetYr 243  
   ||| :|||:||||| | | | | | | | | | | | |  
721 GATGGATATATATATTAACAGCAGCTCTTTGGAGATTAAGATTTCAGGATTA 770  
   ||| :|||:||||| | | | | | | | | | | | |
```

244 AspOlyTyrSerThrAsnSerAspSerMetValIleLysPheGlnAspAs 260
 771 CAATCCCAATCTGATGGGAAATTTATCTAAGGAATAATGATGACCA 820
 260 pAsnProThrAsnSerSerGlnTyrAsnLeuTyrLysIleGly...GlyT 276
 821 CCAAGAAATTCGATATACCTTGTCTGACCTTCTGCGGGGTAAGAATT 870
 276 hrcLulysLeuProTyrAlaValSerLeuMetGlyGlyLysIlePhe 292
 871 ACTCCCAAAATGGAACGTCATTAATATTGCTGACGACGACTTCTGTGA 920
 293 TyrProValasnGlyInsPheThrIleAsnAspSerSerValLeuG 309
 921 AACAATCTGAATGAAATTCACCTGTCACCATCCGAAATCAGTCTC 970
 309 uhrAsnTrpAsnArgValThrAlaValAlaMetProGluValAsnValP 326
 971 CGGTGTGTGTCGCTGACGAGCTTTCGCAATTCGCAAAAGTGAAAT 1020
 326 roValLeuGlyTrpProAlaArgLeuLeuAsnAlaAspValAsnAla 342
 1021 CCCGAGGCTGGCAATATATGGGTAATATTAACTTACACCCAG 1070
 343 ProAspAlaGlyGlnTyrSerGlyGlnIleTyrIleThrPheThrProse 359
 1071 TAGTCAAACTC 1083
 359 rValGluAsnLeu 363
 seq_name: p1r2:S57937

seq_documentation_block:
 CcdT protein precursor - Escherichia coli
 C:Species: Escherichia coli
 C:Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 10-Dec-1999
 C:Accession: S57937
 R:Freihlich, B.J.; Karakashian, A.; Meisen, L.R.; Scott, J.R.
 A:Description: The genes for CS2 pill of enterotoxigenic Escherichia coli and their inte
 A:Reference number: S57934
 A:Accession: S57937
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1364 <FRO>
 A:Cross-references: EMBL:Z47800; NID:q897725; PIDN:CAAB7763.1; PID:q897729
 C:Genetics:
 A:Start codon: TTG
 C:Superfamily: Escherichia colonizing factor antigen cfaE

alignment_scores:
 Quality: 930.50 Length: 366
 Ratio: 3.209 Gaps: 3
 Percent Similarity: 79.235 Percent Identity: 48.361

alignment_block:
 US-09-839-894-9 x S57937 ..

Align seg 1/1 to: S57937 from: 1 to: 364

```

1 ATGAATAAGATTATATATTTTACATGTTCTTCTCAGTACTTTT 50
| | | | | | | | | | | | | | | | | | | | | | | | | |
1 MetLysLysValIlePheValLeuSerMetPheLeuSerGlnValT 17
51 TACATTTCGTATCGGACAGATAAATTCGCGAGATGAAGCATACATA 100
| | | | | | | | | | | | | | | | | | | | | | | | | |
17 rGluGlnSerTrpHisThrAsnValGluAlaGlySerIleAsnLysThr 34
101 ATATTTTGGCCGCGTACAGGAAAC.....GAATCTTCCCAACAA 144
| | | | | | | | | | | | | | | | | | | | | | | | | |
34 InsrIleGlyProIleAspArgSerAlaIleAsnSerTyrProAlaHis 50

```

```

145 AATATATTAATAACCATATTACAGCATACGATGAAGTCACTCTGTA 194
| | | | | | | | | | | | | | | | | | | | | | | | | |
51 TyrIlePheHisGluHisValAlaGlyTyrAsnLysAspHisSerLeuP 67
195 TGTATAGCATGACCTTTTATGTTGCTCTCCACAAATACATTAAGAG 244
| | | | | | | | | | | | | | | | | | | | | | | | | |
67 eAspArgMetThrPheLeuGlySerMetSerSerThrAspAlaSerLys 84
245 CATGTCCAAACCATGAGATTCCTAGCAGTTTCATCGGTGACCGTGA 294
| | | | | | | | | | | | | | | | | | | | | | | | | |
84 LacysProThrGlyGluAsnSerLysSerSer.....GlnGlyGluThr 98
295 AATATTAACATTAACATTTACGAAAAAGAGATTATATAAAGAGACG 344
| | | | | | | | | | | | | | | | | | | | | | | | | |
99 AsnIleLysLeuIlePheThrGluLysLysSerLeuAlaArgLysThr 115
345 ACAATTTAAAGCTATTAACAAATATTGTTCAAAAGCTTAACTGC.... 390
| | | | | | | | | | | | | | | | | | | | | | | | | |
115 uAsnLeuLysGlyTyrLysArgPheLeuTyrGlnSerAspArgCysIleH 132
391 ....CCATCCGGCCTAACCTTAACCTCAGCTCATTTTAACGTATTA 435
| | | | | | | | | | | | | | | | | | | | | | | | | |
132 IstYrValAspLysMetAsnLeuAsnSerHisThrValLysCysValGly 148
436 AAGCGGCTCAGGTGCAAGTTATATTATTTATTCCTGCTGGCGACG 485
| | | | | | | | | | | | | | | | | | | | | | | | | |
149 SerPheThrArgGlyValAspPheThrLeuTyrIleProGlnGlyGlu 165
486 AAAAATTTGCTTTGGTGTATCTGGATGCTACTCTGAAGTTAAGAG 535
| | | | | | | | | | | | | | | | | | | | | | | | | |
165 eAspGlyLeuLeuThrGlyLysIleTrpGluAlaThrLeuGlnLeuArg 182
536 TAAAGACATATAGTACAGCTATGAGAACTTACACTAATATATAC 585
| | | | | | | | | | | | | | | | | | | | | | | | | |
182 allysArgHisTyrAspTyrAsnHisGlyThrTyrLysValAsnIleThr 198
586 ATTAAATTAAGTATTAAGGAAATATTCAGATATGTTACTCAGTTCA 635
| | | | | | | | | | | | | | | | | | | | | | | | | |
199 ValAspLeuThrAspLysGlyAsnIleGlnValTrpThrProLysPheH 215
636 AAGTGACGCTCGGCGTATCTTAACCTTGCCTCAACGTGGGGGCACAT 685
| | | | | | | | | | | | | | | | | | | | | | | | | |
215 sSerAspProArgIleAspLeuAsnLeuArgProGlnGlyAsnGlyLys 232
686 ATATTGGAAGAAATTCGTTGATATGCTTTATGATGATATATGACT 735
| | | | | | | | | | | | | | | | | | | | | | | | | |
232 YrSerGlySerAsnValLeuGlnMetCysLeuTyrAspGlyTyrSerThr 248
736 AACAGCAGCTCTTGGAGATTAAGATTTTCAGATTAACAAATCTGA 785
| | | | | | | | | | | | | | | | | | | | | | | | | |
249 HisSerGlnSerIleGluMetArgPheGlnAspAspSerGlnThrGlyAs 265
786 TGGGAATTTTATCTAAGGAATAAATATGATGACACCAAGAAATTCGAT 835
| | | | | | | | | | | | | | | | | | | | | | | | | |
265 nasGlnTyrAsnLeuIleLysThrGlyGluProLeuLysLysLeuPro 282
836 ATACTTGTCACTTCTTGGCGGGTAAAGTTAACTCCACAACATGGA 885
| | | | | | | | | | | | | | | | | | | | | | | | | |
282 YrLysLeuSerLeuLeuGlnGlyArgGlnPheTyrProAsnAsnGly 298
886 ACGTCATTAATATATGCTGACGACGCTTCGGAACAAACATCGAATAG 935
| | | | | | | | | | | | | | | | | | | | | | | | | |
299 GluAlaPheThrIleAsnAspTrpHisSerLeuPheIleAsnTrpAsn 315
936 AATTACAGCTGTCACCATGCCAGAAATCAAGTTCGCGTGTGTTGGC 985
| | | | | | | | | | | | | | | | | | | | | | | | | |
315 GlnIleLysSerValSerLeuProGlnIleSerIleProValLeuCysTrp 332
986 CTGAGCGTTTGCATTGGATGACAAAAGTGAATCCGAGGCTGGACAA 1035
| | | | | | | | | | | | | | | | | | | | | | | | | |
332 roLysAsnLeuThrPheMetSerGlnLeuAsnAsnProGlnAlaGlyGlu 348
1036 TATATGGTATATTAATGTTACTTTCACACCAAGTAGTCAAACTC 1083

```

```

|||||
349 TyrSerGlyIleuAsnValThrPheThrProSerSerSerIeu 364
seq_name: pir2:AE0541

```

```

seq_documentation_block:
  Probable fimbrial protein tcfD [imported] - Salmonella enterica subsp. enterica serovar
  C:Species: Salmonella enterica subsp. enterica serovar Typhi
  A:Note: This species has also been called Salmonella typhi
  C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
  C:Accession: AE0541
  R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
  R.; T.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
  S.; Moule, S.; O'Gaora, P.
  Nature 413, 848-852, 2001
  A:Authors: Party, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
  A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
  A:Reference number: AB0502; PMID:11677608
  A:Accession: AE0541
  A:Status: Preliminary
  A:Molecule type: DNA
  A:Residues: 1-359 <PAR>
  A:Cross-references: GB:AL513382; PIDN:CAD08773.1; PID:916501589; GSPDB:GN00176
  C:Genetics:
  A:Gene: tcfD

```

```

alignment_scores:
  Quality: 275.50      Length: 303
  Ratio: 1.514        Gaps: 11
  Percent Similarity: 60.066      Percent Identity: 29.703

```

```

alignment_block:
  US-09-839-894-9 x AE0541

```

```

Align seg 1/1 to: AE0541 from: 1 to: 359

```

```

208 TTTTATGTTGTTCTTCACATACCTTAATGAGACATGTCACACG 257
    |||||
79 TrpValCysArgSerAsnArgAsnGluAsnGluGlyAlaCysGluGlu 95
    |||||
258 TCGAATCTCAGCAGTTCAGGTCAGCGGTGAACCAATATATACATTAC 307
    |||||
95 rHsIeuValAlrPTrPTrAlaPheGlyAlaTyrSerIysIleArgIeuA 112
    |||||
308 AATTACGGAAGAAAGAGATTATATAAAGAGAGCTCAACAAATAAAGC 357
    |||||
112 rPheArgGluGlnIleSerHisAlaGlnIleThrIeu..... 124
    |||||
358 TATTAACATATATGTTCAAAAGCTGTAC.....TGGCCATCGGCGCT 401
    |||||
125 .....IleIeuLeuGlySerValArgAspAlaCysTyrThrGlyVa 138
    |||||
402 AACACTTAACCTCAGCTCATTTTAACTGTAATAAAGCGGGCTCA.... 447
    |||||
138 I.....IleAsnMetAsnAlaAlaAlaAlaCysG 147
    |||||
448 .....GTCGCAAGTTATATATTTATATATTCCTGCGGCACTAAATAA 492
    |||||
147 IntPglYArgSerIeuLysIleuArgIleProSerGluGluLeuAlaLys 163
    |||||
493 TTGCCTTTGGTGTATCTGGATGCTACTGTGAAGTTA...AGAGTAA 539
    |||||
164 IleProThrSerGlyThrTrpLysAlaThrIleValIleuAspTyrIeuG 180
    |||||
540 AAGACGATATAGTGAGACCTATGGAACCTACATATAAATATCATATTA 589
    |||||
180 nTrpGlyGlyAspAspProIeuGlyThrSerThrTrpAspIleThrIeuA 197
    |||||
590 AATTACCGAT.....AAGGAAATATTCAGATGAGTTAAGTCACTGTC 633
    |||||
197 snValIthrAspHisPheAlaGluAsnAlaAlaIleTyrPheProGlnPhe 213

```

```

634 ...AAAGTACGCTGCGTCGATCTTAACCTTGCCTCAACCTGTGGGG 680
    |||||
214 GlyThrAlaThrProArgValAspIleAsnIleHisArgMetAsnIle 230
    |||||
681 CACATATATTGGAAGAAATTCTGTGATATGTCCTTTATGATGATATA 730
    |||||
230 rGlnMetSerGlyArgAlaAsnIleuAspMetCysIleuTyrAspGly...G 246
    |||||
731 GTACTAAGCAGCAGCTTTGGAGATAGATTAAGATTTCAGATTAACCT 780
    |||||
246 IyAllyAlaIaArgSerIeuGlnMetLysIleGluIuIySerAsnLysSer 262
    |||||
781 TCGATGAGGAATTTATCTAAGAAATTAATGATGACACCAAGAAAT 830
    |||||
263 GlyThrGlyPheGlnValIleLysSerAspSerAlaAspThr..... 277
    |||||
831 TGCATATATTCTTGTCACTTCTCTTGGCGGTAAAGTTAACTCCACAA 880
    |||||
277 eAspTyrAlaValSerMetAsnTyrGlyArgSerIleProValThrA 294
    |||||
881 ATGGAACGTCATTAATATTGTCGACGAGCTTCTGGAACCAACTGG 930
    |||||
294 rGlyIyAlaGlnPheSerIeuAspAsnValAspLysAlaAlaThrArg... 309
    |||||
931 AATGAAATTACAGCTGTCAACCCAGCAATTCAGTTCCTGCTGTGTG 980
    |||||
310 .....ProValValIeuProGlyGlnArgGlnAlaValArgCys 322
    |||||
981 TTGGCTGTGACGTTTGCATTTGCATGCAAAA.....GTGGAATTC 1021
    |||||
322 sValProValProIeuThrIeuThrGlnProPheAsnIleArgGlu 339
    |||||
1022 CCGAGCGTGGACAAATATATGCGTAATATTAAGTTCACTTCCACACCA 1071
    |||||
339 ysArgSerGlyGluTyrGlnGlyThrIeuThrValThrIeuMetGly 355
    |||||
1072 AGTCAACA 1080
    |||||
356 ThrGlnThr 358

```

```

seq_name: pir2:G90975
seq_documentation_block:
  Probable factor [imported] - Escherichia coli (strain 0157:H7, substrain RIMD 0509952)
  C:Species: Escherichia coli
  C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
  C:Accession: G90975
  R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.
  gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
  DNA Res. 8, 11-22, 2001
  A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli 0157:H7
  A:Reference number: A96929; MUID:21156231; PMID:11258796
  A:Accession: G90975
  A:Status: Preliminary
  A:Molecule type: DNA
  A:Residues: 1-1335 <NAV>
  A:Cross-references: GB:BA000007; PIDN:BA836198.1; PID:913362243; GSPDB:GN00154
  C:Genetics:
  A:Experimental source: strain 0157:H7, substrain RIMD 0509952
  A:Gene: Ecs2775

```

```

alignment_scores:
  Quality: 120.00      Length: 392
  Ratio: 0.628        Gaps: 15
  Percent Similarity: 48.724      Percent Identity: 20.408

```

```

alignment_block:
  US-09-839-894-9 x G90975

```

```

Align seg 1/1 to: G90975 from: 1 to: 1335

```

```

37 TCCTCAGTACTTTTACATTGCTGTATCGGACGATAAATCCCGAGA 86

```

```

||||:  :: |||:|||||:  ::  ::  ::
746 SerIalysIlealatrLeuSerAlaSerAsnGlyValLeuAlaAs 762
87 TGAAGCATTAATAATTTTGGCCCGCGTGACGAGAAATCTCC 136
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::
762 nGluAsnAlaIaAsnThrValSerValaAsnValAlaAspGluLysr 778
137 CCAAACTAATATATTAATACCATTAATACAGCATAC.....AGT 177
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::
779 .....AsnProIleAsnAspHisThrValThrPheAlaValLeuSer 792
178 GAAAGTCATCTGTATGATAGATGACATTTTATGTTGTCTTCA 227
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::
793 GlySerAlaThrSerPheAsnAsnGlnAsn.....ThrAlaLys 805
228 CAATACACTTAATGAGCATGTCACACCATGAGAAATCTAGCAT 277
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
805 sThrAspValaAsnGlyLeuAla...ThrPheAspLeuLysSerLysG 821
278 CGGTACAGCGGTGAACAAATATTAACATTAATTAACGAAAGAAGT 327
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
821 InGluAspAsnThrValGluValThrLeuGluAsnGlyValLysGlnThr 837
328 TTAATA..... 333
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::
838 LeuIleValSerPheValGlyAspSerSerThrAlaGluValAspLeuG 854
334 .....AAAGAGAGCTACAATTAAGCTAT..AAACATTAATGCT 373
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::
854 nLysSerLysAsnGluValAlaAlaAspGlyAsnAspSerAlaThrMet 871
374 TCAAACTGTAACCTGCCATCGCGCTCAACATTAATTAACATGCTAT 423
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
871 hTrAlaThrValArgAspAlaLysGlyAsnLeuAsnAspValLysVal 887
424 AACTGTAATAAAACGCGCTCGAGTGAAGTTATTAATTAATTC 473
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
888 ThrPheAsnValaAsnSerAlaAlaLysLeuSerGlnThrGluVala 904
474 TCGTGGCGAACTAAATAATTTGCTTTGGTGTGATCGGAGCTCATC 523
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
904 nSerHisAsp.....GlyIleAlaThrAlaThrL 914
524 TGAAGTTAAGTAAAGACATATAGTGAACCTTGAACCTTAAC 573
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::
914 euvhr.....SerLeuLysAsnGlyAspTyrThr 923
574 ATAAATTCATAATTAATTAACGATAAGGAAATTTGAGATATGTT 623
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
924 ValThrAlaSerValSerSerGlySerGlnAlaAsnGlnValIlePh 940
624 ACCTCACTCAAAAGTGCAGCTCGCGTGCATTAACCTCGTCAAGT 673
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
940 eIleGlyAspGlnSerThrAlaIleuThrLeuSerVal...ProSerG 956
674 GTGGGGCCACATATATGGAAGAAATCTGTGATATGCTTTATGAT 723
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::
956 LyspIleThrVal..... 960
724 GGATATGATCTTAACAGCAGCTTTGGAG.....ATAAGATT 761
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::
961 .....ThrAsnThrAlaProLeuHisMetThrAlaThrLeuGlnAs 974
762 TCAGATTAACAATCTTAATCTGATGGAAATTTTATCAAGAAATTA 811
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
974 pLysAsnGlyAsnProLeuLysAspLysGluIleThrPheSerValPro 991
812 ATGATGACACCAAGAAATGCATATCTTGTGCATCTTCTTGGCGGT 861
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::
991 snAsp.....ValAlaSerArgPheSerIleSerAsnSerGly 1003
862 AAAAGTTAATCCAAACAAATGAGAGCTCATTA..... 894
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::

```

```

1004 LysGlyMetThrAspSerAsnGlyThrAlaIleAlaSerLeuThrGly 1020
895 .....AATATG 901
1020 rLeuAlaGlyThrHisMetIleThrAlaArgLeuAlaAsnSerAsnVal 1037
902 CTGACGAGCTTCTCTGAAACAAACTGAAATAGAAATTAACGTGAC 951
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
1037 eAspThrGlnProMetThrPheValAlaAspLysAspArgAlaValVal 1053
952 ATGCCAGAAATCAGTGTCCGGTGTGTGTCGCTGCA.....CGTT 995
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
1054 ValLeuGlnThrSerLysAlaGluIleGlyAsnGlyValaAspGlu 1070
996 GCATTTGATGCAAAAGTGAATAATCCGAGCGCTGACAAATATG 1045
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::
1070 rThrLeuThrAlaThrValLysAspPro...PheAspAsnValaLys 1086
1046 ATATTAATGTTACTTTCACACCAAGT 1071
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::
1086 snLeuSerValValPheArgThrSer 1094
seq_name: p1r2:E85822
seq_documentation_block:
probable_invasin 23135 [imported] - Escherichia coli (strain O157:H7, substrain EDL93
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: E85822
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
flier, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoustis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: AB5480; M01D:21074935; PMID:11206531
A:Accession: E85822
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2660 <SNO>
A:Cross-references: GB:AE005174; NID:q12516151; PIDN:AA657041.1; GSPDB:GN00145; UMGCP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: 23135
alignment_scores:
Quality: 120.00 Length: 392
Ratio: 0.628 Gaps: 15
Percent Similarity: 48.724 Percent Identity: 20.408
alignment_block:
US-09-839-894-9 x E85822 ..
Align seg 1/1 to: E85822 from: 1 to: 2660
37 TCTTCAGTACTTTTACATTTGCTATCGGACATTAATTCGCGAGA 86
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::
737 SerAlaLysIleAlaThrLeuSerAlaSerAsnGlyValLeuAlaAs 753
87 TGAAGCATTAATAATTTTGGCCCGCGTGACGAGAAATCTTC 136
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::
753 nGluAsnAlaIaAsnThrValSerValaAsnValAlaAspGluLysr 769
137 CCAAACTAATATATTAATACCATTAATACAGCATAC.....AGT 177
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::
770 .....AsnProIleAsnAspHisThrValThrPheAlaValLeuSer 783
178 GAAAGTCATCTGTATGATAGATGACATTTTATGTTGTCTTCA 227
   |||:  ::  ::  ::  ::  ::  ::  ::  ::  ::
784 GlySerAlaThrSerPheAsnAsnGlnAsn.....ThrAlaLys 796
228 CAATACACTTAATGAGCATGTCACACCATGAGAAATCTAGCAT 277
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
796 sThrAspValaAsnGlyLeuAla...ThrPheAspLeuLysSerLysG 812

```



```
772 AAT...CCTAATCTGATGGAAATTTCGTACAGCAATAAATGA 815  
||| ||| :::::::::::  
326 AsnLeuGIYThAlSerAlalysSerSerpHeIISeTThrThrTh 342  
|| |:::|| |:::::  
816 TGACACCAGAAGAAATT.....GCATATACTTTGTCA..... 846  
|| |:::|| |:::::  
342 rAspleuTrISerIleasnThrSerAlatYSerThrglySerIlSeT 359  
|||||:::|| |:::::  
847 .....CTTCTC 852  
  
359 hValGlutHrGLySarnGrThrThrsErGluallleSerHisValVa 375.  
:::  
853 TGGGGGGGTAAAAGTTTAACTCCACAATAATGAGTCGTAAAAATTC 902  
|||:::|| |:::::  
376 ThrTrISerThLysLeuSerProThrAlatrThrSerLeuThrIea 392  
|||:::|| |:::::  
903 TGACCGACGCTTCCTGGAACAACATCGAAT..... 933  
|:::|:::|| |:::::  
392 acInTrserIletyrSerThrAspSerasnlerThrValGIYThraSp 409  
|:::|:::|| |:::::  
934 .....AGAATT 939  
||  
409 lEhISThrThrsErGluallleSeraspValglutHrIlesErarglu 425  
|||  
940 ACAGCTGCCCATCGATGCCAGCAATCAGTGTCGCGGTGTGTGGCCCTGG 989  
|||:::|| |:::::  
426 ThrAlasErThrValValAlalaPrOthrSerThrThgLYTpThrcI 442  
|||  
990 ACGTTTG 996  
|:::  
442 yLabel 444  
  
seq_name: pir2:D64962
```

seq_documentation_block:

probable membrane protein b1978 - Escherichia coli

C:Species: Escherichia coli

C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 02-Feb-2001

C:Accession: D64962

R:Blauner,F.R.; Plunkett III,G.; Bloch,C.A.; Perna,N.T.; Burland,V.; Riley,M.:; Science 277, 1453-1462, 1997

A:title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: D64962

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-2383 <BLAST>

A:Cross-references: GB:AE000289; GB:U00096; NID:g1788285; PIDN:AAC75042.1; PID:g1788285

A:Experimental source: strain K-12, substrain MG1655

C:Keywords: nucleotide binding; P-loop; transmembrane protein

F:54-70/Domain: transmembrane #status predicted <TM>

F:1564-1571/Region: nucleotide-binding motif A (P-loop)

alignment_scores:

	Quality:	Length:
	109.00	390
	Ratio: 0.565	Gaps: 15
Percent Similarity:	49.487	Percent Identity: 20.000

alignment_block:

US-09-839-894-9 x D64962 ..

Align seg 1/1 to: D64962 from: 1 to: 2383

```
37 TCCTAGACTTTTACATTTCTGATCGCACATAAATTCGGCGACA 86  
|||||:::|| |:::::  
762 SerAlalylsilealThrLeuSerAlaserAnasnnglValleAlaAs 778  
87 TGAAGCACTAATPATTTTTGGCCCCGGGTACAGGAAGCAATCTGCC 136  
:||||:::|| |:::::  
778 molUsnaAlaIaaSnThrValSerValasnValAlaAspgLUIGlySer. 794
```



```
964 .....AGTGTCCGGTGTG..... 978
      ::::: |||||
1257 ILeGlnGluSerIleLysAsnGlyAspLeuThrIleGluValLeuAsnAs 1273
979 .....TGTGGCCCTGGACGTTTGC 997
1273 nProAsnSerAlaSerAnthrIlePheAsnIleAlaProGluLeuTyrA 1290
998 AATTGGATGCAAAAGTGAATAATCC..... 1023
      ::::: |||||
1290 snTyrGlnAlaSerLysGlnAsnProThrGlyTyrSerTyrAspTyrSer 1306
1024 .....GAGCGTGACAAATAT.....ATGGTAATATTAATGTTACTTT 1061
      ::::: |||||
1307 AspAsnGlnAlaGlyThrTyrTyrLeuThrSerAsnIleLysGlyLeuPh 1323
1062 CACACCA..AGTAGTCACAACA 1080
      |||||
1323 eThrProLysGlySerGlnThr 1330
```

seq_name: p1r2:A64904

```
seq_documentation_block:
probable fimbrial protein b1502 - Escherichia coli
C.Species: Escherichia coli
C.Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 29-Sep-1999
C.Accession: A64904
R.Blattner, F.R.: Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Sho, Y.
Science 277, 1453-1462, 1997
A.Title: The complete genome sequence of Escherichia coli K-12.
A.Reference number: A64720; MUID:97426617
A.Accession: A64904
A.Status: nucleic acid sequence not shown; translation not shown
A.Molecule type: DNA
A.Residues: 1-304 <BLAT>
A.Cross-references: GB:AE00247; GB:U00096; NID:g1787773; PIDN:AC74575.1; PID:g1787779;
A.Experimental source: strain K-12, substrain MG1655
C.Superfamily: fimbrial protein fimb
C.Keywords: fimbria
```

```
alignment_scores:
      Quality: 107.00      Length: 357
      Ratio: 0.629      Gaps: 19
Percent Similarity: 47.619      Percent Identity: 21.008
```

alignment_block:

US-09-839-894-9 x A64904 ..

Align seg 1/1 to: A64904 from: 1 to: 304

```
7 AAGATTTTATTT...ATTTTACATGTTGTTTCTCTCAGTACTTTTAC 53
      |||||
8 LysValLeuPheGlyIleTyrLeuLeuMetAlaGlyLysValAlaPheAl 24
54 ATTGCTGATCGGAGATAAATTCGCCGAGATGAAGACATA..... 96
      |||||
24 apheSerCysAsnValAsp.....GlyGlySerSerIleGlyAlaG 38
97 .....ACTAATATTTTGGCCCGTCGACGAAACGAATCTTCCCCAAA 141
      |||||
38 LyrThrIleSerValTyrValAsnLeuAsp...ProValIleGlnProGly 53
142 CATTAATATA.....TTAAATACCATATTACAGCATACAGTGAAG 182
      |||||
54 GlnAsnLeuValAlaSerLeuSerGlnIleIleSerCysTyrAsnAspTy 70
183 TCATACGTCGTATGATGATGACATTTTTATGTTGTCTTCAC...A 229
      |||||
70 rGlyGlyTyrPtyrAsp.....ThrAspHisIleA 80
```

```
230 ATACACTTAATGAGCATGTCTCAACAGCATGAGAAATCTACAGCTCATCG 279
      |||||
80 snLeuValGlnGly.....SerAla 86
280 GTACGCGGTCAACAAATATATACATTACATTACGAAAAAAGAGTTT 329
      |||||
87 PheAlaGly..... 89
330 AATAAAAAGAGCTCAAAATTAAGCTATAAA...CAATTATGTGTCA 376
      ::::: |||||
90 .....SerLeuGlnSerTyrLysGlySerLeuTyrTyrP 101
377 AAAGTTTAATCTGCCATCCGCGCTACACCTTAACCTACGCTCATTTTAC 426
      ::::: |||||
101 snAsnValThrTyrProPheProLeuThrThrAsnThrAsnValLeuAsp 117
427 TGTAAATAAACCGCGCTTACGATGCAAGTTTATTTATATTTCTGCG 476
      ::::: |||||
118 ILeGlyAspLysThrProMetProLeuProLeuLysLeuTyrIle..... 132
477 TGGCGAACTAAAAAATTTGCCCTTTGGTGTATCTGGAGATGCTGTGA 526
133 .....ThrProValGlyAlaAlaGlyValValIleI 144
527 AG.....TTAAGATAAAAGACATATAGTGAACCTAT 561
      |||||
144 ysnIaGlyGluValIleAlaArgIleHisMetLysIleAlaThrLeu 160
562 GGAACT.....TACACTATTAATATACATTTTAA..... 591
      |||||
161 GlySerGlyAsnProArgAsnPheThrPheAsnIleIleSerAsnAsn 177
592 .....TTAACTGATTAAGGAATAATATATCAG 616
177 nValValMetProThrGlyGlyCysThrValAspSerArgAsnValThy 194
617 TATGTTTACCTCAGTTCAAAAGTACGCTCGCATCTTAATCTTGCGT 666
      |||||
194 alAspLeuProAspPheProGlySerAlaGlnIleProLeu..... 207
667 CCAACTGCTGGGGCACATATATTGGAAGAAATTCGTTGATGTCGCTT 716
      |||||
208 .....GlyValTyrCysSer...SerGlnGlnLysLeuSerPh 219
717 TTATGATGATATAGTACATACAGACGCTCTTGAGATATAGATTTACG 766
      |||||
219 eTyrLeuSerGlyAlaThrThrAspSerArgGlnVal...PheAla 235
767 ATAAACAATCCTTAATCTGATGGGAAATTTTATCTAAGAAAAATTAAT 816
      |||||
235 snThrAlaPro.....Asp 239
817 GACACCAAGAAATTCATATATCTTGTCTCTCTGCGCGGTAA... 864
      |||||
240 AlaThrLysAlaSerGlyValGlyValThrLeuMetArgAsnIlyLys11 256
865 .....AGTTTAACCTCCAAATGGAACGTCAT 892
      |||||
256 eLeuAlaThrGlyGlnAsnValSerLeuGlyThrValAsnLysSerLys 273
893 TAAATATTTGCTGACGAGCTTCTGTGAAACAACTGGAATAGATTAACA 942
      ::::: |||||
273 alProLeuGlyLeuSerAlaThrTyrGlyGlnThrGlyAsnLysValSer 289
943 GCTGTACCATGCCAGAAATC 963
      |||||
290 AlaGlyThrValGlnSerVal 296
```

seq_name: p1r2:C90892

```
seq_documentation_block:
probable adhesin [imported] - Escherichia coli (strain O157:H7, substrain RIMD 050995
C.Species: Escherichia coli
```

C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
 C:Accession: C90892
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
 gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genc
 A:Reference number: A99629; MUID:21156231; PMID:11258796
 A:Accession: C90892
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-304 <HAY>
 A:Cross-references: GB:BA000007; PIDN:BAH35530.1; PID:G13361573; GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain RMD 0509952
 C:Genetics:
 A:Gene: Ecs2107
 C:Superfamily: fimbrial protein fimH

alignment_scores:
 Quality: 107.00 Length: 357
 Ratio: 0.629 Gaps: 19
 Percent Similarity: 47.619 Percent Identity: 21.008

alignment_block:
 US-09-839-894-9 x C90892 ..

Align seg 1/1 to: C90892 from: 1 to: 304

```

7  AAGATTTTATTT...ATTTTACATGTTTCTCTGAGTACTTTTAC 53
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8  LysValLeuPheGlyIleTyrLeuLeuMetAlaGlyLysValAla 24
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
54 ATTGCTGATCGGAGATTAATAATCCCGGAGATGAAGCATA..... 96
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24 A:heserysasnValasp.....GlyLysSerIleGlyAlaG 38
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
97 .....ACTAATATTTTGGCCCGCGTGACAGAACGATCTTCCCAAA 141
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
38 LyrThrSerValTyrValasnLeuasp...ProValIleGlnProGly 53
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
142 CATATATATA.....TTAAATACCATATTACAGATACAGTGAAG 182
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
54 GlnAsnLeuValValaspLeuSerGlnHisIleSerCysTrpAsnAspTly 70
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
183 TCATCTCTGTATGATAGATGACTTTTATGTTGTTCTCTC...A 229
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
70 rGlyIlyTyrPyrasp.....ThrsPHisIleA 80
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
230 ATACACTTAATGAGCATGTCCACACAGTGAATCTACAGATCATCG 279
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
80 snLeuValGlnGly.....SerAla 86
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
280 GTCAGCGGTGAACAATATAACATTACAAATTTCGAAAAAGAGATT 329
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
87 PheAlaGly..... 89
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
330 AATAAAGAGAGCTACAAATTAAGGCTATRAA...CAATATTGTCTCA 376
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
90 .....SerLeuGlnSerTyrIlysgLysSerLeuTyrTyr 101
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
377 AAGAGTTTAAGTCCCATCGGCTACACACTTAACATGATCATTTTAC 426
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
101 snAsnValThrTyrProPheProLeuThrThrAsnThrAsnValLeuasp 117
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
427 TGTATAATAAAGCGGCTTCAGGTCAAGTTTATATATATATCTCGC 476
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
118 IleGlyAspLysThrProMetProLeuProLeuLysLeuTyrIle.... 132
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
477 TGGCAACTAAAAAATTTGCTTTGGTGATATCTGGATGCTACTCTGA 526
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
133 .....ThrProValGlyAlaAlaGlyLysValValIleIle 144
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
527 AG.....TTAAGAGTAAAAAGACGATATAGTACGACCTAT 561

```

```

144 ysAlaGlyGluValIleAlaIArgIleHisMetTyrLysIleAlaIArgLeu 160
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
562 GGAAGT.....TACCTATTAATATCACTATTAA..... 591
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
161 GlySerGlyAsnProArgAsnPheThrTrpAsnIleIleSerAsnAsn 177
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
592 .....TTACGTATAGGGAATATATCA 616
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
177 rValValMetProThrGlyGlyCysThrValAspSerAlaGlnValTrv 194
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
617 TATGTTTACCTCAGTTCAAAAGTGACGCTCGCTGATCACTTAATTCGCT 666
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
194 AlAsnLeuProAspPheProGlySerAlaGluIleProLeu..... 207
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
667 CCAACTGGTGGGCGCATATATTGGAAGAAATTTCTGTATATGCTT 716
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
208 .....GlyValTyrCysSer...SerGlnGlnLysLeuSerPh 219
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
717 TTATGATGATATAGTACTTAACAGCAGCTTTGGAGATTAAGATTTCAG 766
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
219 eryLeuSerGlyThrThrThrAspSerAlaArgGlnVal...PheAla 235
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
767 ATAACATCTTAATCTGATGGAATTTTATCTAAGGAAATAAATGAT 816
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
235 snThrAlaPro.....Asp 239
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
817 GACACCAAGAAATTCATATACCTTTGTCACCTCTCTGGCGGTTAA.. 864
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
240 AlaThrLysAlaSerGlyValGlyValSerLeuMetArgAsnGlyLysI 256
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
865 .....AGTTAATCTCAACAATGAAGACGTCAT 892
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
256 eLeuAlaThrGlyGluAsnValSerLeuGlyThrValAsnLysSerLysV 273
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
273 alProLeuGlyLeuSerAlaThrTyrGlyGlnThrGlyAsnLysValSer 289
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
943 GCTGTCACCATGCCAGAAATC 963
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
290 AlaGlyThrValGlnSerVal 296
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

seq_name: p1r2:F85725

seq_documentation_block:
 probable adhesin, fimH type protein Z2206 [imported] - *Escherichia coli* (stra 57:
 C:Species: *Escherichia coli*
 C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: F85725
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
 Iller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
 A:Reference number: AB5480; MUID:21074935; PMID:11206551
 A:Accession: F85725
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-304 <STO>
 A:Cross-references: GB:AE005174; NID:G12515169; PIDN:AAG56266.1; GSPDB:GN00145; UWGP:
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: Z2206
 C:Superfamily: fimbrial protein fimH

alignment_scores:
 Quality: 107.00 Length: 357
 Ratio: 0.629 Gaps: 19
 Percent Similarity: 47.619 Percent Identity: 21.008

alignment_block:
 US-09-839-894-9 x F85725 ..

Align seg 1/1 to: F85725 from: 1 to: 304

```

7 AAGATTTATTT...ATTTTACATGTTTTCTCTCAGTACTTTTAC 53
  |||:|||||  |||:|||||:|||||:|||||:|||||:
8 LysValLeuPheGlyIleTyrLeuLeuMetValGlyValPheAl 24
54 ATTTGCTGTATCGGCAGATAAATTCGGAGATGAAGCAT... 96
  |||:|||||  |||:|||||:|||||:|||||:
24 aPhSerCysAsnValasp.....GlyGlySerSerIleGlyLag 38
97 ....ACTAATATTTTGGCCCGCTGACAGGAAGCATCTCCCCAAA 141
  |||:|||||  |||:|||||:|||||:|||||:
38 LyrThrSerValTyrValAsnLeuasp...ProValIleGlnProGly 53
142 CATATATTA.....TTAATAACCATATTTACATACATGAAG 182
  |||:|||||  |||:|||||:|||||:|||||:
54 GlnAsnLeuValValAspLeuSerGlnHisIleSerCysTrpAsnAspTy 70
183 TCATACTCTGTATGATAGATGACTTTTATGTTGTCTCTCAC...A 229
  |||:|||||  |||:|||||:|||||:|||||:
70 rGlyGlyTrpTyTasp.....ThraSphHisIleA 80
230 ATACACTTAATGAGCATGTCACACAGATGAGATCCTAGCATTCATCG 279
  |||:|||||  |||:|||||:|||||:|||||:
80 snLeuValGlnGly.....SerAla 86
280 GTCAGCGGTAAACAATATACATTACATTACGGAAGAAAGACGTTT 329
  |||:|||||  |||:|||||:|||||:|||||:
87 PheAlaGly..... 89
330 AATTAAGAGAGAGCTACAAATTAAGCCTATAA...CAATATTTGTCA 376
  |||:|||||  |||:|||||:|||||:|||||:
90 .....SerLeuGlnSerTyrTyrGlySerLeuTyrTrpA 101
377 AAGGTGTTACTGCCCATCCGGCTTAACACTTAACGATCATTAAAC 426
  |||:|||||  |||:|||||:|||||:|||||:
101 snAsnValThrTyrProPheProLeuThrAsnThrAsnValLeuasp 117
427 TGTATTAATAACGCGGCTTCAGTGCAGATTTATATATATATTCCTGC 476
  |||:|||||  |||:|||||:|||||:|||||:
118 IleGlyAspLysThrPrometProLeuProLeuLysLeuTyrIle.... 132
477 TGGGAACTAATAAATTTGCTTTGGGTGATCTGGAGTCTACTCGA 526
  |||:|||||  |||:|||||:|||||:|||||:
133 .....ThrProValGlyAlaIleGlyValValIleL 144
527 AG.....TTAAGAGTAATAAGAGCATATAGTGAAGCTTAT 561
  |||:|||||  |||:|||||:|||||:|||||:
144 ysaIaGlyGluValIleAlaTrgIleHisMetTyrLysIleAlaThrLeu 160
562 GGAAGT.....TACATATTAATTCACATATTAA..... 591
  |||:|||||  |||:|||||:|||||:|||||:
161 GlySerLysAsnProArgAsnPheThrTrpAsnIleIleSerAsnAsnse 177
592 .....TTAAGTATTAAGGAATATTCAGA 616
  |||:|||||  |||:|||||:|||||:|||||:
177 rValValMetProThrGlyGlyCysThrValAspSerArgAsnValThrV 194
617 TATGCTTACTCAGTTCAAAAGTACGCGCTCGCATCTTAAGTTCGCT 666
  |||:|||||  |||:|||||:|||||:|||||:
194 alsnLeuProAspPheProGlySerAlaGluIleProLeu..... 207
667 CCAAGCTGGGGGACATATATGTAAGAATCTGTGATATGCTT 716
  |||:|||||  |||:|||||:|||||:|||||:
208 .....GlyValTyrCysSer...SerGlnGlnLysLeuSerPh 219
717 TTAATGATGATATAGTACTAAGACAGCTCTTGAGATAGATTCAGG 766
  |||:|||||  |||:|||||:|||||:|||||:
219 eTyrLeuSerGlyThrThrThrAspSerAlaArgGlnVal...PheAla 235
767 ATAACAATCTTAATCTGATGGAAATTTATCTAAGAAAATTAATGAT 816
  |||:|||||  |||:|||||:|||||:|||||:

```

```

235 snThrAlaPro.....Asp 239
817 GACACCAAGAAATTCATATACCTTGTCTACTCTCTTGGCGGTAA... 864
  |||:|||||  |||:|||||:|||||:|||||:
240 AlaThrLysAlaSerGlyValGlyValSerLeuMetLysAsnGlyLysI 256
865 .....AGTTTAAGTTCACCAATGAAGCTCAT 892
  |||:|||||  |||:|||||:|||||:|||||:
256 eleuAlaThrGlyGluAsnValSerLeuGlyThrValAsnLysSerLysV 273
893 TAAATATTCGTGACGACGCTTCTGTGAACAAACATGATATGATACA 942
  |||:|||||  |||:|||||:|||||:|||||:
273 alProLeuGlyLeuSerAlaThrTyrGlyGlnThrGlyAsnLysValSer 289
943 GCTGTACCATGCCAGCAATC 963
  |||:|||||  |||:|||||:|||||:|||||:
290 AlaGlyThrValGlnSerVal 296

seq_name: p1r2:A38109

seq_documentation_block:
autolysin - Enterococcus faecalis
C:Species: Enterococcus faecalis
C:date: 24-Jul-1992 #sequence_revision 24-Jul-1992 #text_change 15-Oct-1999
C:Accession: A38109
R:Bellevue, C.; Potvin, C.; Trudel, J.; Asselin, A.; Bellemare, G.
J. Bacteriol. 173, 5619-5623, 1991
A:Title: Cloning, sequencing, and expression in Escherichia coli of a Streptococcus f
A:Accession: A38109; MUID:91358349
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-671 <BEL>
A:Cross-references: GB:M58002; NID:9153658; PIDN:AAA67325.1; PID:9829194

alignment_scores:
Quality: 106.00 Length: 309
Ratio: 0.716 Gaps: 17
Percent Similarity: 47.896 Percent Identity: 23.625

alignment_block:
US-09-839-894-9 x A38109 ..
Align seg 1/1 to: A38109 from: 1 to: 671

121 AGGAGCAATCTTCCCAACATATATA...TTAATAACCATATTAC 167
  |||:|||||  |||:|||||:|||||:|||||:
309 ArgTyrAlaThrAspProSerTyrAsnAlaLysLeuAsnValIleTh 325
168 AGCATACAGTGAAGTCACTCTGTATGATAGATGATTTTATGTT 217
  |||:|||||  |||:|||||:|||||:|||||:
325 rAlaTyr.....AsnLeuThrGlnTyrAspThrProSerSerGlyLys 340
218 TGTCTCTCACATACACTTAATGAGATGTCACACAGTGAAGATCT 267
  |||:|||||  |||:|||||:|||||:|||||:
340 snThrGlyGlyGlyThrValAsnProGlyThrGlyGlySerAsnGln 356
268 AGCAGTTCATCGGTC.....AGCGGTGAACA.....AA 296
  |||:|||||  |||:|||||:|||||:|||||:
357 SerGlyThrAsnThrTyrTyrThrValLysSerGlyAspThrLeuAsn 373
297 TATTAACATTAACAATTT.....ACGAAATAAGAGTTTA.... 330
  |||:|||||  |||:|||||:|||||:|||||:
373 sIleAlaIleGlnTyrGlyValSerValAlaAsnLeuArgSerTrpAsn 390
331 ..ATAAAGAGAGCTACAAATTAAGGCTATTAACAATTAATTTTCA 378
  |||:|||||  |||:|||||:|||||:|||||:
390 LysLeuSerGlyAspLeuIlePheValGlyGlnLysLeuIleValLys 406
379 AGGTATACGCGCCATCGGCTTACACTTAAGTCACTGATTTTAACT 428
  |||:|||||  |||:|||||:|||||:|||||:
407 GlyAla.....SerGlyAsnThrGlyGlySerLysGlnGlyGly 420

```

```

429 TAATAAAACGGCGCTTCAGGTCAGATTATTAATATATCTGCTG 478
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
430 TAAATAAAT...GlnSerGlyThrAsnThrTyrThrValLysSerG 436
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
436 LysAspThrLeuAsn..... 440
529 TTAAGACTAAAGACGATATAGACACTGAGAACTTACTACTATAA 578
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
441 .....LysIleAlaIleAlaGlnTyrGly..... 447
579 TATCACTATTAAATTAAGTATAGTAAAGAAATATCAGATATGTTACCTC 628
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
448 ValThrVal.....AlaAsnLeuArgSerTrp..... 456
629 AGTCAAAAGTACGCTGCGCTGATCTTAACTTG..... 663
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
457 .....AsnGlyIleSerGlyAspLeuIlePheValGlyGlnLysLeu 470
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
664 .....CGTCCACTGCTGGGGGACATATATTTGAAAGAAATCTGT 704
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
471 IleValLysLysGlyThrSerGlyAsnThr..... 480
705 TGATATGCTTTATATGATGATATATCTACTAAGACAGCTCTTGGAGA 754
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
481 .....GlyIleSerSerAsnGlySer..... 488
755 TAAGATTTCAGATACATCTTAATCTGATGGAAATTTATCTAAG 804
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
489 .....AsnAsnAsnGlnSerGlyThrAsnThrTyrThrIle 501
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
805 AAAATAATATGATGACACCAAGAAATGCA.....TATACTTTGTCA.. 846
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
502 LysSerGlyAspThrLeuAsnLysIleAlaIleGlnTyrGlyValSerVa 518
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
847 .....CTTCTCTTGGCGG 859
518 IAlaAsnLeuArgSerTrpAsnGlyIleSerGlyAspLeuIlePheVal 535
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
860 GTAATAATTAATCTCAACAAATGACGATTAATATGCTGACGCA 909
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
535 LysGlnLysIleIleValLysGlyThrSerGlyAsnThrGlySer 551
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
910 GCTTCTCTGGAACAACTGGAATAGA 936
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
552 SerAsnGlySerAsnAsnAsnGln 560

```

seq_name: p1r2:AC2507

seq_documentation_block:
 hypothetical protein al17235 [imported] - Anabaena sp. (strain PCC 7120) plasmid pCC7120
 C:Species: Anabaena sp.
 A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
 C:Accession: AC2507
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuriitz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
 A:Reference number: AB1807; M01D:21595285; PMID:11759840
 A:Accession: AC2507
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-843 <KUR>
 A:Cross-references: GB:BA000020: PIDN:BA078319.1; PID:q17135773; GSPDB:GN00180
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: al17235
 A:Genome: plasmid

alignment_scores:
 Quality: 105.50 Length: 376
 Ratio: 0.567 Gaps: 18
 Percent Similarity: 49.468 Percent Identity: 18.883

alignment_block:
 us-09-839-894-9 x AC2507 ..

Align seg 1/1 to: AC2507 from: 1 to: 843

```

229 AATCACTTAATGAGCATGTCACACAGTACAGAT..... 264
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
297 AsnIleLeuAspGlySerLeuValPheThrGlnAsnHISGlyPheLysTh 313
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
265 .....CCTACAGTTCATCGCTACGCGGTGAACAA 295
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
313 rGlyAlaValLysIleAspAlaGlnSerLeuAsnIleGlnLysSerVa 330
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
296 ATATACATTACAA.....TTTACG..... 315
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
330 snLeuAlaLeuSerAlaIleTyrThrSerAsnPheGlyPheThrProGly 346
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
316 GAAATAAGAACTTAATTAATAAGAGAGCTACAAATTAAGCTTAATAACA 365
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
347 GluSerIleGlnLeuAspValLysAspValThrIleGlnGlyGlnIle 363
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
366 ATATTTGTTCAAAAGCTTAATCGCCATCCGCGCTA...ACACTTAAC 412
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
363 eAlaThrThrThrPheThrAsnAlaProSerGlyLeuIleThrIleAsn 380
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
413 CAGCTCATTTAAC.....TGTAATATAAAC 438
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
380 eAsnSerLeuLysIleSerGlyAspThrProSerTyrAlaAsnProAsp 396
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
439 GCGGCTTCAGCTGCAAGTTTATTAATTAATTCCT..... 474
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
397 GlyLeuGlyGlyIleAsnThrPheSerTyrSerSerGlyLysGlyLys 413
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
475 .....GCTGGCGAACTAAATAATTTGCTTTTGT.....GGTATCTGGG 514
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
413 pIleAlaGlyLysIleAsnAsnIleIleIleGlyLeuAspLysAlaPheA 430
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
515 AT.....GCTACTCGAAGTTAAGATA 537
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
430 snThrValAlaSerGlySerGlyAlaGlyAsnLeuPheLeuLeu 446
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
538 AAAAGACGATATAGTACACTTATGAACTTAACTAATATACATAT 587
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
447 GluAsnLeuIleIleLysAspGlyAlaIleSerLeuGlySerSerThrI 463
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
588 TAAATTAACGTATAGGAAATATTCAGATATGCTTACCTCAGTTCAAA 637
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
463 eArgSerGlyGlnGlyLysAsnVal.....PheIleL 474
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
638 GTGACGCTCGGTGATCTTAC.....TTCCGTCACACT 672
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
474 ySerGlnAsnIleAspIleSerGlyGlnSerAlaLeuLeuArgProSer 490
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
673 .....GCTGGGCGACATATTTGGAAG 695
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
491 AsnIleThrSerSerThrPheGlyHisGlyAsnLysGlyAsnIleAspI 507
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
696 AAATTCGTTGATATGCTTTTATGATGATATAGTAAACAGACACT 745
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
507 eAsnThrLeuAsnLeuIleIleSerAsnGlyGlyLysSerSerSer 524
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
746 CTTTGAG..... 753
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
524 hrLeuSerAlaGlyLysAlaGlyAsnIleSerIleAsnSerSerAsnSer 540
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
754 .....ATATGATTTACAGATTAACATCTTAA.....TC 782
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
541 IleAsnValValGlyThrAsnIleAsnSerAsnSerProSerPheIleAs 557

```

```
783 TGATGGGAATTTATCTAAGAAAAATAATGATGACCAACCAAGAAATG 832
      ::::::::::||| ||| :::: ::::
557 nserSerAsnPheLeu.....ValAspProAsnLeuGlnLysLeu 572
      ::|
833 CATATACTTGTCACTTCTCTGGCGGTAAAGT..... 867
      ||| ||||| |||:::
572 eutYrArgInProProLeuLeuIleGlyGlnIaGlyAsnIlePheLeu 588
      ::::::::::TTAAGTCCACAAATGSAAGTCATTAAT..... 897
868 ..... 897
589 AsnThrAspIleIleAsnIleSerAsnGlyGlyLeuIleAsnIaArgAs 605
      :::::::::::::::||||| ::::
898 .....ATTGCTGACGACGCTTCTCTGGAACAACTGGAATAGAATTA 940
      ::| |||||::: :::: ||| |||
605 nGluGlyValAsnAspIaGlyAsnIleArgIleSerAlaAsnThrIleA 622
      :: :::: |||::: ::
941 CAGCTGTCACCATGCCAGAAATCAGTGTCCGGTGTGTGGCT... 987
      :: :::: |||::: ::
622 snIleAsnSerGlnGlyGluValAsnIaThrThrIleGlyGluGly 638
      ::|::: |||:::
988 GGACGTTTGCATTTGATGCCAAAGTGGAATCCCGAGGCTGACATA 1037
      |||::: |||:::
639 GlyAsnIleIleLeuAsnSerArg.....AsnLeuPn 649
      ::|::: |||:::
1038 TATGGTAATATTATGTTACTTTCACA 1065
      ::|::: |||:::
649 eleuAsnAsnSerArgIleThrIaThr 658
```


OM of: US-09-839-894-9 to: SwissProt_40.* out_format : pfs
Date: Jul 1, 2002 11:23 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+22p.model -DEV=xlp
-O=/cgn2.1/USPROT/US09839894/runat.01072002.111123.17552/app-query.fasta.1.1160
-DB=SwissProt_40 -OFMT=fasta -SUFFIX=isp -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-XGAPOP=6.000 -XGAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-DELO=6.000 -DELEXT=7.000 -START=1 -MATRIX=blsnum62
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALLEN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09839894 @CGNL.1.73 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPPX -WAIT -THREADS=1

Search information block:

Query: US-09-839-894-9
Query length: 1086
Database: SwissProt_40.*
Database sequences: 105224
Database length: 38719550
Search time (sec): 35.410000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
SwissProt_40:CFAB_ECOLI + 1558.50	2359.16	2.96-124	360	1	P25734 Escherichia coli, cfa
SwissProt_40:SAGI_YEAST + 114.00	160.20	0.0487	650	1	P20840 saccharomyces cerevis
SwissProt_40:GASJ_YEAST + 108.00	152.93	0.1534	524	1	Q03655 saccharomyces cerevis
SwissProt_40:DECO_ECOLI + 107.00	156.06	0.1768	304	1	P77886 escherichia coli, hly
SwissProt_40:ALYS_EFMFA + 106.00	147.77	0.2319	671	1	P37710 enterococcus faecalis
SwissProt_40:IGAL_HAEIN + 99.00	128.47	1.00	1849	1	P45386 haemophilus influenza
SwissProt_40:IDJ3_MERJA + 98.50	137.23	0.9900	608	1	Q58788 methanococcus jannasch
SwissProt_40:SP21_YEAST + 98.00	134.58	1.11	758	1	P35209 saccharomyces cerevis
SwissProt_40:NL33_YEAST + 98.00	130.96	1.16	1157	1	P36161 saccharomyces cerevis
SwissProt_40:CDG2_PABMA + 95.50	131.31	1.80	713	1	P30735 paenibacillus macerans
SwissProt_40:RF1M_YEAST + 95.00	135.22	1.89	413	1	P30785 saccharomyces cerevis
SwissProt_40:WAPA_BACSU + 95.00	120.41	2.23	2334	1	Q07833 bacillus subtilis, w
SwissProt_40:NI00_YEAST + 94.50	127.25	2.26	959	1	Q02629 saccharomyces cerevis
SwissProt_40:LEU2_BOCR + 94.00	132.96	2.32	471	1	P48573 buchnera aphidicola
SwissProt_40:THBG_SHEEP + 93.50	132.11	2.52	412	1	P50450 ovis aries (sheep), b
SwissProt_40:MURE_BUCAT + 93.00	123.30	3.08	1167	1	P56956 bacillus thuringiens
SwissProt_40:CLAA_BACTU + 93.00	119.67	3.21	1783	1	Q49460 mycoplasma genitalis
SwissProt_40:CYF_CVACA + 92.50	129.07	3.15	320	1	Q91544 cyanidium caldarium
SwissProt_40:AGMI_CANAL + 92.50	123.35	3.36	544	1	Q9P4V2 candida albicans (yea
SwissProt_40:OAR_MYXNA + 92.50	116.82	3.62	1061	1	P38756 myxococcus xanthus, c
SwissProt_40:EABI_YEAST + 92.50	127.44	3.83	2278	1	P53714 saccharomyces cerevis
SwissProt_40:CFAC_ECOLI + 91.50	113.54	4.01	869	1	P25734 escherichia coli, cfa
SwissProt_40:POLG_HETIM + 91.50	115.62	4.38	2193	1	Q06479 human enterovirus 71
SwissProt_40:ELTB_CLOPE + 90.50	122.08	4.40	309	1	P01585 clostridium perfringens
SwissProt_40:NCMA_XENLA + 90.50	120.66	4.94	863	1	P30664 xenopus laevis (afri
SwissProt_40:ENRK_HUMAN + 90.50	118.39	5.07	1019	1	P98073 homo sapiens (human)
SwissProt_40:HUS2_SCHPO + 89.50	118.39	6.06	1328	1	Q08811 schizosaccharomyces
SwissProt_40:NGA2_YEAST + 89.00	125.09	6.14	1113	1	P40578 saccharomyces cerevis
SwissProt_40:ANJ3_MOUSE + 88.00	128.27	7.08	270	1	Q9ZM56 heliobacter pylori
SwissProt_40:CYB_CHIRE + 88.00	125.28	7.32	381	1	P23663 chlamydomonas reinhard
SwissProt_40:CYB_CHEP + 88.00	125.28	7.32	381	1	P23663 chlamydomonas reinhard
SwissProt_40:VGC_HSEVB + 88.00	123.52	7.46	468	1	P12889 equine herpesvirus ty
SwissProt_40:HAP_HAEIN + 87.50	114.18	8.30	1394	1	P45387 haemophilus influenza
SwissProt_40:UBPA_YEAST + 87.50	113.86	8.66	792	1	P53874 saccharomyces cerevis
SwissProt_40:GPRY_MOUSE + 87.00	118.42	9.44	711	1	Q98Y22 Shope fibroma virus
SwissProt_40:ETP2_SVKA + 87.00	113.33	10.00	1290	1	P18640 clostridium botulinu

SwissProt_40:THBG_BOVIN + 86.50 122.35 9.87 411 1 Q9TC36 bos taurus (bovine
SwissProt_40:VEL_HPV24 + 86.50 118.47 10.31 647 1 P36725 human papillomaviru
SwissProt_40:SPHR_AMPV + 86.50 114.73 10.76 1002 1 P29815 ameba moorei en
SwissProt_40:WEB1_YEAST + 86.50 112.68 11.01 1273 1 P38968 saccharomyces cer
SwissProt_40:RW1_DROME + 86.50 111.05 11.22 1540 1 Q9V7H4 drosophila melanog

seq_name: SwissProt_40:CFAB_ECOLI

seq_documentation_block:

ID	CFAB_ECOLI	STANDARD	PRT	360 AA
AC	P25734			
DT	01-MAY-1992 (Rel. 22, Created)			
DT	01-MAY-1992 (Rel. 22, Last sequence update)			
DE	1-AUG-1992 (Rel. 23, Last annotation update)			
DE	CFA/I fimbrial subunit E (Colonization factor antigen I subunit E).			
GN	CFAB.			
OS	Escherichia coli.			
OG	Plasmid NTP53.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Escherichia			
OX	NCBI_TaxID=562;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=9232981; PubMed=1352712;			
RA	Jordi B.A.M., Willshaw G.A., van der Zelfst B.A.M., Gastra W.;			
RT	"The complete nucleotide sequence of region 1 of the CFA/I fimbrial			
RT	operon of human enterotoxigenic Escherichia coli.";			
RL	DNA Seq. 2:257-263(1992).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
DR	EMBL; M55661; AAC41417.1; -			
KW	Antigen; Fimbria; Plasmid.			
SQ	SEQUENCE 360 AA; 39903 MW; 691509B63ABE69CE CRC64;			

alignment_scores:

Quality: 1558.50 Length: 361
Ratio: 4.504 Gaps: 1
Percent Similarity: 95.845 Percent Identity: 81.163

alignment_block:

US-09-839-894-9 x CFAB_ECOLI ..
Align seg 1/1 to: CFAB_ECOLI from: 1 to: 360

1 ATGAATTAAGATTATTTATTTATTTTACATTTGTTTCTCTCACTTTT 50
1 Metanlystleleuphellepetherleuphethleuphethleupheth 17
51 TACATTGCGTATCGGAGATTAATTCGCGGATGGAATGAACATACATA 100
17 ethrhealavalseralalaspplysasnproglyserglutansmettr 34
101 ATATTTTGGCCCGGCGTACAGAACAGATCTTTCCCAACATATATA 150
34 snthrlleglyprohlsasparglyglyserseerproiletrynsnle 50

```

151 TTAATAACATATTACAGCATACGAGAAAGTCATCTGTATGATAG 200
|||||
51 LeuSerSerTyrLeuThrAlaTyrAsnGlySerHisHisLeuTyrAspAr 67
201 GATGACCTTTTATGTTGCTCTCCACAAATACATTAATGAGCATGTC 250
|||||
67 gmeSerPheLeuSerSerSerGlnSerHisLeuAsnGlyAlaCysP 84
251 CAACAGTGAGAAATCTAGCAGTTCATCGTCAGCGGTGAAACAAATATA 300
|||||
84 roSerSerAspAlaProGlyThrAlaThrIleAspGlyGluThrSer 100
301 ACATTACATTTACGGAAGAAAGATTAAATAAAGAGACATCAAT 350
|||||
101 ThrLeuGlnPheThrGluLysArgSerLeuIleLysArgGluLeuGln 117
351 TAAAGGCTATAACAAATTATGTTCAAAAGTTCATGCGCCATCGGCC 400
|||||
117 elySolYtyrLysGlnPheLeuPheLysAsnAlaAsnGlyProSer 134
401 TAACACTTAACCTAGCTCATTTTAACCTTAATAAAGCGCGCTTCAGAT 450
|||||
134 euAlaLeuAsnSerSerHisPheGlnCysAsnArgGluGlnAlaSer 150
451 CCAAGCTTATATTTATATATATCTCGTCGCGCAACTAAAAAATTTGCC 500
|||||
151 AlaThrLeuSerLeuTyrIleProAlaGlyGluLeuAsnLysLeuPro 167
501 TGGTGAGTCTCGGAGTCTACTCTGAACCTTAAGATAAAGACATATA 550
|||||
167 egLYglYAlaItrPAsnAlaValLeuLysLeuAsnValLysArgA 184
551 GTGAGACCTATGGAACCTTACACTATATATACATTAATTAATTAAT 600
::: |||||
184 splrThrTyrGlyThrTyrThrIleAsnIleThrValAsnLeuThr 200
601 AAGGAAATATTCAGATATGCTTACCTCAGTTCAAAGTGACGCTCGCT 650
|||||
201 LysGlyAsnIleGlnIleTyrPheProGlnPheLysSerAsnAlaArg 217
651 CGATCTTACTGCTGCTCACTGCTGGGGGACATATATTTGGAAGAATT 700
|||||
217 lAsPheAsnLeuArgProThrGlyGlyThrTyrIleGlyArgAsn 234
701 CTGTGATATGCTGCTTATGATGATATGATATGATATGATATGAT 750
|||||
234 ervAlaPheMetCysPheTyrAspGlyTyrSerThrAsnSerSerLeu 250
751 GAGATAGATTTAGAGATTAACAATCTTAATCTGATGGGAATTTATCT 800
|||||
251 GluIleArgPheGlnAspAsnSerLysSerAspGlyLysPheTyrLe 267
801 AAGGAAATTAATATGATGACACCAAGAAATGCAATATCTTGTACCTC 850
|||||
267 ulysLysIleAsnAspAspSerLysGluLeuValTyrThrLeuSer 284
851 TCTTGGCGGTAAAGTTAACTCCACAAATGGAACGTCTTAATATAT 900
|||||
284 euleAlaIleGlyLysAsnLeuThrProThrAsnGlyGlnAlaLeuAs 300
901 GCTGACGACGCTCTCTGGAAGAAACATGGAATAGATATGACGCTG 950
|||||
301 ...AsnThrAlaSerLeuGluThrAsnTrpAsnArgIleThrAlaVal 316
951 CATGCCGAATTCAGTGTCTCGGCTTGTGCTGGCTGGAGCTTTGCAT 1000
|||||
316 rmeProGluIleSerValProValLeuCysTrpProGlyArgLeuGln 1050
1001 TGGATGCAAAAGTGAAGAAATCCGAGCTGGAACAATATATGGAATAT 1050
|||||
333 euAspAlaLysValLysAsnProGluAlaGlyGlnTyrMetClyAsn 349
1051 AATGTTACTTTCACACCAAGTAGTCACAACTC 1083

```

```

seq_name: SwissProt_40:SAGL_YEAST
350 LysIleThrPheThrProSerSerGlnThrLeu 360
:::|||||
seq_documentation_block:
ID SAGL_YEAST STANDARD; PRT; 650 AA.
AC P20840;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alpha-agglutinin precursor (Ag-alpha-1).
GN SAG1 OR AGAL1 OR YJR004C OR J1418.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90014768; PubMed=2677666;
RA Lipke P.N., Wojciechowicz D., Kurjan J.;
RT "AG alpha 1 is the structural gene for the Saccharomyces cerevisiae
RT alpha-agglutinin, a cell surface glycoprotein involved in cell-cell
RT interactions during mating."
RL Mol. Cell. Biol. 9:3155-3165(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C;
RC MEDLINE=90005993; PubMed=2676603;
RA Hauser K., Tanner W.;
RT "Purification of the inducible alpha-agglutinin of S. cerevisiae and
RT molecular cloning of the gene."
RL FEBS Lett. 255:290-294(1989).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;
RA de Haan M., Smits P.H.M., Grivell L.A.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
RN [4]
RP PARTIAL SEQUENCE, CARBOHYDRATE-LINKAGE SITES, AND DISULFIDE BONDS.
RX MEDLINE=96064684; PubMed=7592821;
RA Chen M.-H., Shen Z.-M., Bobin S., Kahn P.C., Lipke P.N.;
RT "Structure of Saccharomyces cerevisiae alpha-agglutinin. Evidence for
RT a yeast cell wall protein with multiple immunoglobulin-like domains
RT with atypical disulfides."
RL J. Biol. Chem. 270:26168-26177(1995).
CC -|- FUNCTION: CELL SURFACE GLYCOPROTEIN PROMOTING CELL-CELL CONTACT
CC TO FACILITATE MATING. (SACCHAROMYCES CEREVISIAE A AND ALPHA CELLS
CC EXPRESS THE COMPLEMENTARY CELL SURFACE GLYCOPROTEINS A-AGGLUTININ
CC AND ALPHA-, RESPECTIVELY, WHICH INTERACT WITH ONE ANOTHER TO
CC PROMOTE CELLULAR AGGREGATION DURING MATING.
CC -|- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC (POSSIBLE).
CC -|- INDUCTION: BY EXPOSITION TO PHEROMONE (A-FACTOR) SECRETED BY THE
CC OPPOSITE MATING TYPE CELLS (TYPE A).
CC -|- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED.
CC -|- SIMILARITY: TO C. ALBICANS ALS1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M28164; AAA34417.1;
DR EMBL; X16861; CAA34752.1;
DR EMBL; X87611; CAA60926.1;
DR EMBL; Z49504; CAA89526.1;
DR PIR; S22835; S22835.
DR SGD; S0003764; SAG1.
KW Glycoprotein; Cell adhesion; Signal; GPI-anchor; Repeat.

```



```

816 TGACACCAAGAAAT.....GCATPACCTTGCA.....846
|||
342 rAspleuThrSerIleasnThrSerAlaYrSerThrGlySerIleSerT359rIleSerT
847 .....CTTC852CTC
359 hValGluThrGlyAsnArgThrThrSerGluValIleSerHisValVal375r
853 TTGGCGGGTAAAGTTAACTCCACAATGGACGTCATTAATATGCG902r
: : : : : ||||| |||||
376 ThrThrSerThrLysLeuSerProThrIaThrThrSerLeuThrIleAl392rIleAl
903 TGACGACGCTCTCTCGTAACAACAACTGGAA.....933rCTCT
| : : : : : |||
392 agInThrSerIleYrSerThrAspSerAsnIleThrValGlyThrAspI409rThrAspI
934 .....AGAAAT939r
409 leHisThrThrSerGluValIleSerAspValGluThrIleSerArgGlu425r
940 ACAGCTGTACCATGCCAGAAATCAAGTCTCCGGGTGTGGTGGCCGTGG989r
||||| ||| : : : ||| ||
426 ThrIaSerThrValValAlaIaIaProThrSerThrThrGlyTrpThrGl442rThrGl
990 ACGTTTG996
| : :
442 yAlaMet444

```

seq_name: SwissProt_40:GAS3_YEAST

```
seq_documentation_block:
ID      GAS3_YEAST      STANDARD;      PRT;      524 AA.
```

DT	01-NOV-1997 (Rel. 35, Created)
DT	01-NOV-1997 (Rel. 35, Last sequence update)
DT	01-MAR-2002 (Rel. 41, Last annotation update)
DE	GAS3 protein precursor.
GN	GAS3 OR YMR215W OR YMR261.09.
OS	Saccharomyces cerevisiae (Baker's yeast).
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX	NCBI_TaxId:4932;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=S288C / AB972;
RA	Deidman K., Brown D., Bowman S., Barrell B.G., Rajandream M.A.,
RA	Walsh S.V.,
RL	Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN	[2]
RP	IDENTIFICATION.
RX	MEDLINE=20529944; PubMed=11079560;
RA	Pardo M., Ward M., Bains S., Molina M., Blackstock W., Gil C.,
RA	Nombela C.,
RT	"A proteomic approach for the study of Saccharomyces cerevisiae cell
RT	wall biogenesis.";
RL	Electrophoresis 21:3396-3410(2000).
CC	-!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC	(Potential).
CC	-!- SIMILARITY: BELONGS TO THE GAS1 FAMILY.
CC	-----
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
EMBL:	Z49809; CAAB9930.1; -.
CC	COMPUENFAST-2DPAGE: Q03655; -.
DR	SGD; S0004828; GAS3.
DR	Glycoprotein; Membrane; GPI-anchor; Signal.
WV	

FT	CHAIN	1	21	POTENTIAL.
FT	SIGNAL	22	?	GAS3 PROTEIN.
FT	PROPEP	?	524	REMOVED IN MATURE FORM.
FT	DOMAIN	458	496	SER-RICH.
FT	CARBOHYD	201	201	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	269	269	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	350	350	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	385	385	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	404	404	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	422	422	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	504	504	N-LINKED (GLCNAC. . .) (POTENTIAL)
SO	SEQUENCE	524 AA;	56793	MM; 3A2E0DBD3E8690 CR664;

[illegible]

alignment_block:
US-09-839-894-9 x GAS3_YEAST

Align seg 1/1 to: GAS3_YEAST | from: 1 to: 524

```

335 GGCTTAAACATT...TTGTCAAGTGTTACGCGCCATCGG 398
      |||||      :|||:|||||
264 GlyTyrAspLysLeuAsnSerThrPheGluSphAlaValIlePro... 278
      |||||      :|||:|||||
399 CCTAACTTAACTCAGCTCATTTTAACTGTAAATAAACGGCGCT... 444
      |||       :|||:|||||

```

279LeuIlePheSerGluTyrGlyCysAsnLysAsnThrProArgT 293

445 ...TCAGTGCAGTTTATATTATATATTCTGCTGGCACTAAA 489

293 hrpAaspgluValserGluGlyLeuTyr.....GlyGlyLeuLys 306

490 AATTGGCCTTTTGGTGTATCTGGGATGCTACTCTGAAGTTAAGAGTAA 539

```

|||:: |||||::|
307 AsnValPheSerGlyGlyLeu.....Va 314

```

540 AAGACGATATAGTGAGACCTATGGAAGTTACACTATAAATATCACTAATTA 589

314 IYRGluTYRThrGluAlaAsnAsnTYRGlyLeu.....ValL 328

590 AATTAACTGATAAGGGAATATTCAGATATGGTTACTCTCAGTTCAAAGT 639

```

328 yslLeuAspSerpGlySerLeu.....ThrTyrLysasp 339
      |||||  |||::|||:::|:::|  ::|||:::

```

640 GAC..GCTGCGTGATCTTAACCTGGT.....CCAACCTGG 674

340 AspPheValAsnLeuGlnSerGlnLeuLysAsnValSerLeuProThrThr 356

675 TGGGGGCACATATATTGGAAGAAATTCTGTGATATGTGC..... 714

356 RLYSGLSERGLUTIESERASPRIETRYLYSCYSPASPSA 373

715TTTATGATGATATAGTACTAACAGCAGCTCTTGAG 753

373 laleThrasnileTyrSerGlyPheGlyTThrasnAsnPhethrLeuPro 389

754ATAGATTTCAGGATAACAATCC 776

390 SerGlnProAlaGluIleAlaAsnMetIleGluTyrGlyValAsnGlyTh 406

777 TAAATCTGATGGAAA..... 792

406 rasnThr...GlylysileLeuthraspsTyralavalProThrThrphea 422

793 .TTTATCTAAGGAAATTAATGATGACACCAAGAATTGCATATACT 840

422 sntYrThrllelYsAsnslYsAspAspThr...IleSerAlaThr 436

841 TTGTCACCTCTCTTGGGGGTA¹AAAGTTTAACTCCACAACAATGGAACGTC 890

[illegible]

```

seq_documentation_block:
ID      VIDEO_ECOLI      STANDARD:      PR:      304 AA.
AC      P7588:
DT      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Hypothetical fimbrial-like protein ydeQ precursor.
GN      ydeQ OR B1502.
OS      Escherichia coli.
OC      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC      Escherichia.
OX      NCBI_TaxId=562;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=K12 / MG1655.
RX      MEDLINE=97426617; PubMed=9278503;
RA      Blattner F.R., Plunkett G. II, Bloch C.A., Perna N.T., Burland V.,
RA      Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA      Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA      Mau B., Siao Y.;
RT      "The complete genome sequence of Escherichia coli K-12.";
RL      Science 277:1453-1474(1997).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=K12;
RX      MEDLINE=97251357; PubMed=9097039;
RA      Alta H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA      Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA      Makino K., Maki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA      Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,
RA      Sampei G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
RA      Takekoshi K., Takeuchi Y., Wada C., Yamamoto Y., Horichi T.;
RT      "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT      corresponding to the 28.0-40.1 min region on the linkage map.";
RL      DNA Res. 3:363-377(1996).
CC      -1 SIMILARITY: TO TYPE-1 FIMBRIAL SUBUNITS.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
CC      EMBL: AE000247; AAC74575.1; -
DR      EMBL: D90792; BAA15175.1; -
DR      EMBL: D90793; BAA15183.1; -
CC      EcoGene: EG13799; ydeQ.
DR      InterPro: IPR000259; Fimbr1a1.
DR      Pfam: PF00419; Fimbr1a1; 1
DR      Hypothetical protein; Fimbr1a1; signal; Complete proteome.

```

[illegible]

```

194 AlaSpleuProAspPheProGlySerAlaGluLeuProLeu..... 207
667 CCAACTGGTGGGACATATATGGAAGAAATTCGTGTATATGTGCT 716
208 .....GlyValTyrCysSer...SerGluGlnLysSerPhe 219
717 TTATGATGATATAGTACTACACAGCTCTTGGAGTAAGATTGACG 766
219 eTylLeuSerGlyAlaThrThrAspSerSerArgGlnVal...PheAla 235
767 ATAACATCCCAATCTGATGGGAATTTATCTAAGAAATTAATGAT 816
235 snrHraAlaPro.....Asp 239
817 GACACCAAGAAATTCATATATCTTGCACTCTTGCGGGTAA... 864
240 AlaThrIysAlaSerGlyValGlyValThrLeuMetArgAsnGlyVal 256
865 .....AGTTTAACCTCAACAATGGAACGTCAT 892
256 eLeuAlaThrGlyGluAsnValSerLeuGlyThrValAsnLysSerIysV 273
893 TAAATATGCTGACGACGCTTCTCTGGAACAACCTGGAATGATTA 942
273 alProLeuGlyLeuSerAlaThrTyrGlyGlnThrGlyAsnLysValSer 289
943 GCCTGACCATGCCAAGATC 963
290 AlaGlyThrValGlnSerVal 296

```

seq_name: SwissProt_40:ALYS_ENTFA

seq_documentation_block:

ID ALYS_ENTFA STANDARD: PRI: 671 AA.

AC P37710:

DT 01-OCT-1994 (rel. 30, Created)

DT 01-OCT-1994 (rel. 30, Last sequence update)

DT 16-OCT-2001 (rel. 40, Last annotation update)

DE Autolysin Precursor (EC 3.2.1.-) (Peptidoglycan hydrolase)

DE (Beta-glycosidase).

OS Enterococcus faecalis (Streptococcus faecalis).

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;

OC Enterococcus.

OX NCBI_TaxID=1351;

RN (1)

RP SEQUENCE FROM N.A.

RX MEDLINE-91358349; PubMed=1679432;

RA Belliveau C., Potvin C., Trudel J., Asselin A., Bellemare G.,

RT Cloning, sequencing, and expression in Escherichia coli of a

RL J. Bacteriol. 173:5619-5623(1991).

CC -1- FUNCTION: HYDROLYZES THE CELL WALL OF E. FAECALIS AND

M. LYSODEICTICUS. MAY PLAY AN IMPORTANT ROLE IN CELL WALL GROWTH

AND CELL SEPARATION.

CC -1- SUBCELLULAR LOCATION: Secreted (Probable).

CC -1- DOMAIN: LYSM REPEATS ARE THOUGHT TO BE INVOLVED IN PEPTIDOGLYCAN

BINDING.

CC -1- SIMILARITY: BELONGS TO FAMILY 73 OF GLYCOSYL HYDROLASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC EMBL: M58002; AAA67325.1; -

DR PIR: A38109; A38109

DR Interpro: IPR002901; Amidase_4.

DR Interpro: IPR002482; LysM.

DR Pfam: PF01832; Amidase_4; 1.

```

DR Pfam: PF01476; LysM; 5.
DR SMART: SM00047; Lys2; 1.
DR SMART: SM00257; LysM; 5.
KW Hydrolase; Glycosidase; Bacteriolytic enzyme; Cell wall;
KW Cell division; Septation; Repeat; Signal.
FT SIGNAL 1 53
FT CHAIN 54 671
FT REPEAT 363 405 LYSM 1.
FT REPEAT 431 473 LYSM 2.
FT REPEAT 499 541 LYSM 3.
FT REPEAT 567 609 LYSM 4.
FT REPEAT 629 671 LYSM 5.
SQ SEQUENCE 671 AA; 70442 MW; 3458268C6C1A4A33 CRC64;

```

alignment_scores: Quality: 106.00 Length: 309
Ratio: 0.716 Gaps: 17
Percent Similarity: 47.896 Percent Identity: 23.625

alignment_block:

US-09-839-894-9 x ALYS_ENTFA

Align seg 1/1 to: ALYS_ENTFA from: 1 to: 671

```

121 AGAAGCAATGCTCCCAAGATATATA...TTAATACCATATTAC 167
||| .....|.....|.....|.....|.....|.....|
309 ArgTyrAlaThrAspProSerTyrAsnAlaLysLeuAsnValIleThr 325
168 AGCATACAGTGGAAGATCATCTGTATGATAGAGTACTTTTATGTT 217
|||.....|.....|.....|.....|.....|.....|
325 AlaTyr.....AsnLeuThrGlnTyrAspThrProSerSerGlyAla 340
218 TGCTTCTCAATACATACATTAATGAGCATGTCACACATGGAATCT 267
.....|.....|.....|.....|.....|.....|
340 snrHrgLysGlyGlyThrValAsnProGlyThrGlyGlySerAsnGln 356
268 AGCAGTTCATCGGTC.....AGCGTGGAACA.....AA 296
|||.....|.....|.....|.....|.....|.....|
357 SerGlyThrAsnThrTyrTyrThrValLysSerGlyAspThrLeuAsn 373
297 TATACATTCATCAATTT.....ACGAAAAAAGATTTA... 330
|||.....|.....|.....|.....|.....|.....|
373 sIleAlaIleGlnTyrGlyValSerValAlaAsnLeuArgSerTyrAsn 390
331 ..ATAAAGAGAGCTACAAATTAAGCTATTAACATATATGTTCAA 378
|||.....|.....|.....|.....|.....|.....|
390 LysLeuSerGlyAspLeuIlePheValGlyGlnLysLeuIleValLysLys 406
379 AGNCTTAACGCGCATCCGCTTAACACTTAACGACGCTATTAACTG 428
|||.....|.....|.....|.....|.....|.....|
407 GlyAla.....SerGlyAsnThrGlySerGlyAsnGlyLys 420
429 TAAATAAAGCGCGCTCAGGTGCAAGTTATATTAATATTCGCGTG 478
|||.....|.....|.....|.....|.....|.....|
420 rAsnAsnAsn...GlnSerGlyThrAsnThrTyrTyrThrValLysSer 436
479 GCGACATAAANAATTCGCTTTGCTGCTATCTGCTACTCTGAAG 528
|||.....|.....|.....|.....|.....|.....|
436 LysAspThrLeuAsn..... 440
529 TTAAGATAAAGAGCATATAGAGACCTATAGAACTTACACTATAA 578
|||.....|.....|.....|.....|.....|.....|
441 .....LysIleAlaIleGlnTyrGly..... 447
579 TATCACTATTAAATTAACGTAAGGAATATTCAGATGCTTACCTC 628
|||.....|.....|.....|.....|.....|.....|
448 ValThrVal.....AlaAsnLeuArgSerTyr..... 456
629 AGTTCAAAAGTACGCTCGGCTGATTAACCTG..... 663
|||.....|.....|.....|.....|.....|.....|
457 .....AsnGlyIleSerGlyAspLeuIlePheValGlyGlnLysLeu 470

```

```

664 .....CGTCCAACTGGTGGGGCCACATATATTGGAAGAAATCTGT 704
      |||:|||||:|||||
471 ILeValLysLysGlyThrSerGlyAsnThr..... 480
      |||:|||||:|||||
705 TGATATGCTTTTATGATGATTAAGTACTAACAGCAGCTCTTGAGAGA 754
      |||:|||||:|||||
481 .....GlyGlySerSerAsnGlySer..... 488
      |||:|||||:|||||
755 TAAGATTTCAGATTAACATCTTAATCTGATGGAAATTTATCTAAG 804
      |||:|||||:|||||
489 .....AsnAsnAsnGlnSerGlyThrAsnThrTyrThrIle 501
      |||:|||||:|||||
805 AAATAAATGATGACACCAAGAAATGCA.....TATACTTTGTCA.. 846
      |||:|||||:|||||
502 LysSerGlyAspThrLeuAsnLysIleAlaIleThrGlyValSerVa 518
      |||:|||||:|||||
847 .....CTTCTCTTGGCGG 859
      |||:|||||:|||||
518 IAlaAsnLeuArgSerThrPasnGlyIleSerGlyAspLeuIlePheAla 535
      |||:|||||:|||||
860 GTAAGACTTAATCACTCAACAAATGACGTCAATTAATATTGTCGAGCA 909
      |||:|||||:|||||
535 LysGlnLysIleIleValLysLysGlyThrSerGlyAsnThrGlySer 551
      |||:|||||:|||||
910 GCTTCTCTGGAAACAACCTGGAATAGA 936
      |||:|||||:|||||
552 SerAsnGlyGlySerAsnAsnAsnGln 560
      |||:|||||:|||||

```

seq_name: SwissProt_40:IGA4_HAEIN

seq_documentation_block:

IGA4_HAEIN STANDARD: PRT: 1849 AA.

AC P45386:

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 01-NOV-1995 (Rel. 32, Last annotation update)

DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).

GN IGA.

OS Haemophilus influenzae.

OC Bacteria: Proteobacteria; gamma subdivision; Pasteurellaceae;

OC Haemophilus.

OX NCBI_Taxid=727;

RP SEQUENCE FROM N.A.

RC STRAIN=NH7 HK61;

RA MEDLINE=92234949; PubMed=1373717;

RA Poulsen K.; Reinholdt J.; Killian M.;

RT "A comparative genetic study of serologically distinct Haemophilus

influenzae type 1 immunoglobulin A1 proteases."

RL J. Bacteriol. 174:2913-2921(1992).

- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A

PRODUCING INTRACT FC AND FAB FRAGMENTS.

- CATALYTIC ACTIVITY: Cleavage of immunoglobulin a molecules at

certain Pro-I-Xaa bonds in the hinge region. No small molecule

substrates are known.

- SUBCELLULAR LOCATION: Secreted.

- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC

SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE

OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE

DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY

SIMILARITY).

- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M87491; AAA24968.1; -
 DR MEROPS; S06.001; -

```

DR InterPro: IPR000710; IGA_S6.
DR Pfam: PF02395; IGA1; 1.
DR PRINTS: PRO0921; IGASERPTASE.
KW Hydroxylase; Serine protease; Transmembrane; Zymogen; Signal.
FT CHAIN 1 25
FT PROPEP 26 1021
FT ACT_SITE 1022 1849
FT ACT_SITE 299 299
SQ SEQUENCE 1849 AA; 202957 MW; 79A7D018C7150A8A CRC64;

```

alignment_scores:

Quality: 99.00 Length: 325
 Ratio: 0.593 Gaps: 20
 Percent Similarity: 51.385 Percent Identity: 23.385

alignment_block:

US-09-839-894-9 x IGA4_HAEIN ..

Align seg 1/1 to: IGA4_HAEIN from: 1 to: 1849

```

55 TTTGCTGATCGCGAGATAAATTCGGAGATGAAGCATTAATAT 104
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
754 PheLysAlaThrThrMetAsnValThrGlyAsnAlaSerLeuTyrSer.. 769
105 TTTTGGCCCGCGTGACAGAGAAATCTCCCAACATTAATATTAA 154
   |||:|||||:|||||:|||||:|||||:|||||:|||||
770 .....GlyArgAsnValAla.....AsnIleThrS 778
155 ATAACCATATTACAGCATACAGTAAAGT.....CATACTCTGTATGAT 198
   |||:|||||:|||||:|||||:|||||:|||||:|||||
778 eArAsn...IleThrAlaSerAsnAsnAlaGlnValHisIleGlyTyrLys 793
199 AGATGACATTTTATGTTGCTTCT..... 225
   |||:|||||:|||||:|||||:|||||:|||||:|||||
794 ThrGlyAspThrValCysValArgSerAspTyrThrGlyTyrValThrCy 810
226 ..CAAAATACA...CTTAATGAGCATGTCACACAGAGAGATCTGACA 271
   |||:|||||:|||||:|||||:|||||:|||||:|||||
810 SHISAsnSerAsnLeuSerGluLysAlaLeuAsnSerPheAsnPro.... 825
272 GTTCATCGCTCAGCGGTGAACAATATTAACA.....TTACAATTTACG 315
   |||:|||||:|||||:|||||:|||||:|||||:|||||
826 ..ThrAsnLeuArgGlyAsnValAsnLeuThrGluAsnAlaSerPheThr 841
316 GAAAAAAGAGCTTAATTAAGAGAGCTACAAATTAAGGCTTAACA 365
   |||:|||||:|||||:|||||:|||||:|||||:|||||
842 LeuGlyLysAlaAsnLeuPheGlyThrIleGlnSerIleGlyThrSerGI 858
366 ATTTATTTTCAAAAGCTTTAACTCCCATCCGCGCTTAACATTAACACG 415
   |||:|||||:|||||:|||||:|||||:|||||:|||||
858 nValAsnLeuLysGlu...AsnSerHisThrPheIleThrIleGlyAsnSer. 873
416 CTCATTTTAACTTAATTA...AACCGCGCTTCAGGTGCAAGTTTAAAT 462
   |||:|||||:|||||:|||||:|||||:|||||:|||||
874 .....AsnValAsnGlnLeuAsnLeuThrAsnGly.....His 884
463 TTATATATTCCTGCTGGCGAACAATAAATTTGCTTTGGGTATCTG 512
   |||:|||||:|||||:|||||:|||||:|||||:|||||
885 IleHisLeuAsnAlaGlnAsnAspAlaAsn..... 894
513 GGATGCTACCTCTGAAGTTAAGATTAAGCAATATAGTGAGACCTATG 562
   |||:|||||:|||||:|||||:|||||:|||||:|||||
895 .....LysValThrThrTyr 900
563 GAATTAACATTAATATATCATTAATTAATTAATTAATTAATTAATTA 612
   |||:|||||:|||||:|||||:|||||:|||||:|||||
900 snThrLeuThrValAsn.....SerLeuSerGlyAsnGlySerPhe 913
613 CAGATATGTTTACCTCAGTTCAAAAGTGACCT...CGCGTCGATCTTAA 659
   |||:|||||:|||||:|||||:|||||:|||||:|||||
914 TyrTyrThrPheValAspPheThrAsnAsnLysSerAsnLysValValAs 930

```

```
660 CTTCGCTCAACTGTGGGGCACATAT.....ATTGAA 694
      |||||::: |||
930 nlusseralaThrGlyasnPherleuGlnValAlaAspLysThrGly 947
      ||::: |||
695 GA...AATTGTGATATGCTTTATGATGATTAAGTACTAACAGC 741
      :: |||
947 lProAsnHIsAsnGlnLeuThrLeuPheAspLaseAsnAlaThrArg 963
      :: |||
742 AGCTCTTGAGATAGATTTCAGGATAACAAT...CCTAAACTGATG 788
      |||||::: |||
964 AsnAsnLeuGlnValThrLeuAlaAsnGlySerValAspArgGlyAlaTr 980
      ||::: |||
789 GAAATTTATCTAAGAAATAATATGATGACACCAAGAAATTCATATA 838
      ||::: |||
980 pLysTyrLysLeuArgAsnValAsnGly....ArgTyrAspLeuTyrA 995
      :: |||
839 CTTTGACACTCTCTTGGCGGTAAAGTTAACTCCACAAATGAGAGC 888
      :: |||
995 snProGlnValGlnLysArgAsnGlnThrValAspThrTrpAsnLleThr 1011
      :: |||
889 TCATTAAAT.....ATTGCTGACGAGCTCTCTGGAACAACACTGGA 932
      :: |||
1012 ThrProAsnSpiLleGlnAlaAspAlaProSerAlaGlnInsAsnAsnGln 1028
      :: |||
933 TAGAATTACAGCTGTCCACCATGCCA 957
      :: |||
1028 uGlnLleAlaArgValGlnLthrPro 1036
      :: |||
seq_name: SwissProt_40:YD93_METUA

seq_documentation_block:
ID YD93_METUA STANDARD; PRT; 608 AA.
AC 058788;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ1393.
GN MJ1393.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN J
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hursb M.A., Kaine B.P., Borodovsky M.,
RA Kleen H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RA jannaschii."
RT Science 273:1058-1073 (1996).
RL -1- SIMILARITY: TO M.JANNASCHII MJ1394 AND A.FULGIDUS AF2028.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U67579; AAB99403.1; -.
DR TIGR: MJ1393; -.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 4 24 POTENTIAL.
SQ SEQUENCE 608 AA; 66768 MW; 010FAF1C29F8C73C CRC64;
```

```
alignment_scores:
      Quality: 98.50
      Ratio: 0.505
      Percent Similarity: 46.539
alignment_block:
US-09-839-894-9 x YD93_METUA
Align seg 1/1 to: YD93_METUA from: 1 to: 608
1 ATGAATTAAGATTTTATTT.....ATTTTACATTTGTTTCTCTTC 41
  ||:::|||||::: |||
1 MetArgLysLeuIlePheMetAlaLeuLeuMetSerLeuPheIleGln 17
42 AGTACTTTTACATTTGCTGATGCGGACGATAAATTCCTC..... 81
  ::|||::: |||
17 YThrValPheGlyTyr.....GlyAspAsnGlyProLeuTyrValA 31
  ::|||::: |||
82 .....GGAGATGAAGC 93
31 LatYrTyrGlnLysTyrAsnLleThrGlyAsnThrThrGlyAspGlyLeu 47
94 ATACTAATATTTTGTGGCCGCGCTGACAGAGCAAGATCTCCCAACAA 143
  ::|||::: |||
48 ValSerSerThrIle.....GlutSerIleThrGlyTyr 58
144 TAATATATTAATTAACCATATTTACAGCATACAGTGAAGATCATCTGT 193
  ::|||::: |||
58 rIleValIleAsnAsn.....ThrGlyThrThrIleAsnAspThrLeu 73
194 ATGATAGGATGACCTTTTATGTTGTTGCTCTCACAATACA..... 234
  |||||::: |||
73 YrAspValTrpValAlaValAsnIleSerAsnAsnIleThrGlyProGln 89
235 .....CTTAATGA.....GCATGTC 251
  ::|||::: |||
90 ValTyrValAsnGlyThrProLysGlyValPheIleGlnSerAlaPro 106
252 AACCACTGAGAAAT...CTTAGCATTCATCG.....G 280
  ::|||::: |||
106 oAlaTyrThrAsnLeuProAsnAlaAsnThrTyrIleHisIleProIle 123
281 TCAGCGGTGAACCAATATTAACATTAATTAACGAAAAAGAGCTTAA 330
  ::|||::: |||
123 eUrProAsnAsnSerTyrValIleIleLysPheAlaIleAspLysSerIle 139
331 ATAAAAAGAGAGCTACAAATTAAGGCTATTAACATTAATGTTCAAG 380
  |||||::: |||
140 ThrGlyValProLeuIleIleAsnGln.....ThyTyrSerAs 152
381 TGTTAACTGCCCATCGGCGCTTAACACTTAAC...TCAGCATTTTAAC 427
  ::|||::: |||
152 pThrIlePheIleProSerIleArgLeuSerAsnTrpSerValTyrLeuAsn 169
428 GTAATTAACCGCGGCTTCAGGTGCAAGTTTATATTATATTCCTGCT 477
  ::|||::: |||
169 lSerArgAsnValSerAla.....LeuProAla 178
478 GGGCAA.....CTAAAAAATTTGCC 497
  ::|||::: |||
179 ThrAspThrProValSerValIleMetThrLysTyrLeuSerAsnSpr 195
498 TTTTGGGTGATCTGGGATCTACTCTGAAGTTAAGAGTAAAAAGCAG 547
  |||
195 O..... 195
548 ATAGTGAGACCTATGAACTTACACT.....ATAATATCACTATT 588
  ::|||::: |||
196 .....AsnAsnTyrGlySerAspThrThrAsnPheLeuAsnIleThrGly 210
  ::|||::: |||
589 AATTAACGTATTAAGCAATATTAATTCAGATATG.....TT 623
  ::|||::: |||
```



```

297 AlaHisLysIleTyrLeuAlaAspArgLysThrGluAlaGlnGlnAsp 313
738 CAGCAGCCTTTGGAGATTAAGATTTCAGATTAACAATCTTAATCTGATG 787
313 nasGlnHisGlnAsnIleAlaTyrGluIleAsnThrLeuIleAsnAspA 330
788 GGAATTTTATCTAAGCAATAATGATGAC...ACCAAGAAATTTGCA 834
330 snThrIleGlnArgThrLysIleAspSerValSerLysArgPheasp 346
835 TATACTTTGCTACTTCTCTTGGCGGTTAAAGTTAACTCCACAATATGC 884
347 PheMetLeuAsnLysArgLysSerThrLysLysValSerProGlyIleAl 363
885 AACG.....TCATTAATATTT 900
363 aThrIleAlaLysLysProAlaSerIleAsnIle 374
seq_name: SwissProt_40:N133_YEAST

```

```

seq_documentation_block:
ID N133_YEAST STANDARD; PRT; 1157 AA.

```

```

AC P36161;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Nucleoporin NUP133 (Nuclear pore protein NUP133).
GN NUP133 OR YKR082M OR YKR402.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JUXR; PubMed=7813444;
RX MEDLINE=95112817; PubMed=7813444;
RA Doye V., Wepf R., Hurt E.C.;
RT "A novel nuclear pore protein Nup133p with distinct roles in poly(A)+
RT RNA transport and nuclear pore distribution.";
RL EMBO J. 13:6062-6075(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94262327; PubMed=8203164;
RA Garcia-Canalejo J., Baladron V., Esteban P.F., Santos M.A., Bou G.,
RA Remacha M.A., Revuelta J.L., Ballesta J.P.G., Jimenez A., del Rey F.;
RT "The complete sequence of an 18,002 bp segment of Saccharomyces
RT cerevisiae chromosome XI contains the HBS1, MRP-L20 and PRP16 genes,
RT and six new open reading frames.";
RL Yeast 10:231-245(1994).
CC -!- FUNCTION: INVOLVED IN POLY(A)+ RNA TRANSPORT AND NUCLEAR PORE
CC DISTRIBUTION.
CC -!- SUBCELLULAR LOCATION: Nuclear pore complex.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: X80066; CAA56372.1; -
DR EMBL: 227116; CAA81633.1; -
DR EMBL: 228307; CAA82161.1; -
DR PIR: S38160; S38160.
DR PIR: S39123; S39123.
DR SGD: S0001790; NUP133.
KW Nuclear protein; Transport; Transmembrane.
FT TRANSMEM 217
FT SIMILAR 98 420 TO NUP120 (AA 434-763).
SQ SEQUENCE 1157 AA; 13319 MW; C8BDBB7D709C5C08 CRC64;

```

alignment_scores:

```

Quality: 98.00 Length: 339
Ratio: 0.587 Gaps: 21
Percent Similarity: 49.263 Percent Identity: 22.124

```

alignment_block:

```

US-09-839-894-9 x N133_YEAST
Align seg 1/1 to: N133_YEAST from: 1 to: 1157

```

```

142 CATATATATATTAATTAACCATATTACAGCATAC..... 174
96 TyrGlyLeuValAsnAspHisLysLysValTyrIleThrAsnIleHisLe 112
175 .AGTGAAGCATACATCTGTGTGATGATGACATCTTT..... 210
112 rThrGlnLysAspThrProTyrIleThrValProPheArgSerAspAspA 129
211 .....TTATGTTGCTCTTCACAAATACACTT 237
129 snAspGluIleAlaValAlaProArgCysIleLeuThrPheProAlaThr 145
238 AATGAGCATGTCCACACGATGAGAAATCCTAGAGTTTCATCGTCAGCG 287
146 MetAspGluSerProLeuAlaLeuAsnProAsnAspGlnAspGluThrGl 162
288 TGAA..... 291
162 yGlyLeuIleIleIleLysGlySerLysAlaIleTyrTyrGluAspIleA 179
292 .....ACAAATATATACATTACAAATTTACGAAAGAAAGATTATATA 333
179 snSerIleAsnAsnLeuAsnHisLysLeuSerGluLysPheSer..... 193
334 AAAGAGAGCTACAAATTA.....GGCTATTAACAATT 368
194 ..HisGluLeuGluLeuProIleAsnSerSerGlyGlyLysCysAs 209
369 ATTGTTCAAAAGTGTACATGCG..CCATCGGCGCTAACATTAACTCA. 414
209 pLeuMet.....LeuAsnCysGluProAlaGlyIleValLeuSerThrA 224
415 .....GCTCATTTTAACCTGTATTAATAAAGCGCGCTTCAGAGTCA 453
224 snMetGlyArgIlePhePheIleThrIleArgAsnSerMetCylLysPro 240
454 AGTTATATATTATATATTCTGTCGGCACTAAATAAATTTGCCCTTTGG 503
241 GlnLeuLysLeu.....GlyLysLeuAsnLysProPheLe 253
504 T...GGTATCGGATGCTACTCTGCACTAAGA.....GTAAAA 541
253 slenGlyIleThrSerLysIlePheAsnThrAsnSerValValSerL 270
542 GAGCATATAGTGAAGACCTAT...GGACTTACATTAATAATACAT 585
270 euAlaGAsnGlyProIleLeuGlyLysGlyThrArgLeuValTyrIleThr 286
586 ATTAATAATTAAGTAAAGAAATTCAGATATGTTACTCAGTTCAA 635
287 .....ThrAsnLysGlyIlePheGlnThrTrp.....GlnLeuSe 298
636 AAGTGAAGCTCGCGCTGATCTTAACCTGCGTCAACAGTGGTGGGCACAT 685
298 Ala.....ThrAsnSerHisProThr..... 305
686 ATATTGAAGAATTCGTGTGATATGTCCTTTTATGATGATATAGTACT 735
306 .....LysLeuIleAspValAsnIleTyrGluAlaIleLeuGlu 318
736 AACAGACGCTCTTTC.....GAGATTAAGATTTCAGGA 767

```

319 SerLeuGlnAspLeuTyProPheAlaHisGlyThrLeuLysIleTyrAs 335
768 TAACAATCTTAATATGAT.....GGAAATTTATCTAAGAAATTA 811
335 PserHisProLeuGlnAspLeuSerGlnLeuPheLeuSerIleT 352
812 ATGATGACACC...AAGAAATTCATATATCTTCACTCTCTGCGC 858
352 YrAspSerSerCysAsnGluThrTyrIleLeuSerThrIleIlePhe 368
859 GGTAAAGTTAACTCCACAAATGCAAGCTCATTAATATGCTGACGC 908
369 AspSerSer.....SerAsnSerPheThrIlePheSerTh 380
909 AGCTTCTCTGGAACAACACTGGAATAGATTACAGCTGTACCATG...C 955
380 TTYrArgLeuAsnThrPheMetGluSerIleThrAspThrLysPheLysP 397
956 CAGAAATCAGTGTCCG 972
397 TolyIlePheIlePro 402

seq_name: SwissProt_40:CDG2_PAEMA

seq_documentation_block:
ID CDG2_PAEMA STANDARD; PRT; 713 AA.
AC P31835;
DT 01-JUL-1993 (rel. 26, Created)
DT 01-JUL-1993 (rel. 26, Last sequence update)
DT 15-DEC-1998 (rel. 37, Last annotation update)
DE Cyclomalto-dextrin glucanotransferase precursor (EC 2.4.1.19)
DE (Cyclodextrin-glycosyltransferase) (CGTase).
OS Paenibacillus macerans (Bacillus macerans).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
NCBI_TaxID=44252;
[1]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 28-37.
RA Sugimoto T., Kubota M., Sakai S.;
RT Polypeptide possessing cyclomalto-dextrin glucanotransferase
activity.
RL Patent number GB2169902, 23-JUL-1986.
CC -1- CATALYTIC ACTIVITY: Degrades starch to cyclodextrins by formation
of a 1,4-alpha-D-glucosidic bond.
CC -1- COFACTOR: BINDS TWO CALCIUM IONS.
CC -1- SUBUNIT: MONOMER.
CC -1- MISCELLANEOUS: CGTASE MAY CONSIST OF TWO PROTEIN DOMAINS: THE ONE
IN THE AMINO-TERMINAL SIDE CLEAVES THE ALPHA-1,4-GUCOSIDIC BOND
IN STARCH, AND THE OTHER IN THE C-TERMINAL SIDE CATALYZES OTHER
ACTIVITIES, INCLUDING THE RECONSTITUTION OF AN
ALPHA-1,4-GUCOSIDIC LINKAGE FOR CYCLIZING THE
MALTOOLIGOSACCHARIDE PRODUCED.
CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
KNOWN AS THE ALPHA-AMYLASE FAMILY.
DR PIR: S26589: ALBSXR.
DR HSSP: P43379: 1CDG.
DR InterPro: IPR000461; Alpha_amylase.
DR InterPro: IPR002044; CBD_4.
DR InterPro: IPR002909; IPT_TIG.
DR Pfam: PF00128; alpha-amylase; 1.
DR Pfam: PF02806; alpha-amylase_C; 1.
DR Pfam: PF00686; CBD_4; 1.
DR Pfam: PF01833; TIG; 1.
DR ProDom: PD001568; CBD_4; 1.
KW transferase; Glycosyltransferase; Calcium; signal.
FT SIGNAL 1 27
FT CHAIN 28 713 CYCLOMALTODEXTRIN GLUCANOTRANSFERASE.
FT DOMAIN 28 165 A1.
FT DOMAIN 166 229 B.
FT DOMAIN 230 434 A2.
FT DOMAIN 435 522 C.
FT DOMAIN 523 609 D.
FT DOMAIN 610 713 E.

FT ACT_SITE 256 256 BY SIMILARITY.
FT ACT_SITE 285 285 BY SIMILARITY.
FT ACT_SITE 356 356 BY SIMILARITY.
SQ SEQUENCE 713 AA; 76857 MW; 5A287BCC4AAFE635 CRC64;

alignment_scores:
Quality: 95.50 Length: 324
Ratio: 0.654 Gaps: 12
Percent Similarity: 45.062 Percent Identity: 20.062

alignment_block:
US-09-839-894-9 x CDG2_PAEMA ..

Align seg 1/1 to: CDG2_PAEMA from: 1 to: 713

202 ATGACTTTTATGTTGCTCTGCACAAATACACTTAATGAGCATGTCC 251
8 LeuThrSerValSerMetSerValGlyIleAlaLeuGlyAlaLeuPr 24
252 AACCACTGAGATCTCTAGCACTTCATCGCGCGGTAAACAATATTA 301
24 OvalTyrPalasPro...AspThrSerValAsnAsnLysLeuAsnPhes 40
302 CATTACATTTTACGAAAGAAAGAGTTAATTAAGAGAGCTACAAAT 351
40 er.....ThrAspThrValTyrGlnIleValThrAspArgPheVal 53
352 AAAGCTTATTAACAATATTTGTCACAAAGGTATTAAGTCCGCGCCT 401
54 AspGly.....AsnSerAlaAsnAsnProThrGlyAl 64
402 AACACTTACCTACGCTCATTTTAACGTGAATTAAGACCGGCTTCAGTG 451
64 aAlapheSerSerAspHisSerAsn..... 72
452 CAAGTTTATTTATATATTCCTGCTGCGAAGCTAAATTTGCTTT 501
73LeuLysLeuTyr.....Phe 77
502 GGTGGTATCTGGGATGCTACTCTGAAGTTAAGAGTAAAGAGCATATAG 551
78 GlyGlyAspTrpGlnGly.....IleThrAsnLysIleAs 89
552 TGAGACCTATGAGACTTACACTATTAATATCATCTTTAATTAAGTATA 601
89 nasPglYtyr.....LeuThrGlym 96
602 AGGGAATATTCAGATATGTTACCTCAGTCAAAAGAGAGCGTGGCGTC 651
96 etGlyIleThrAlaLeuTrpIleSerGlnProValGluAsnIleThra 112
652 GATTTAATCTTCCGTCACAACTGTGGGGGACATATATTGA..... 693
113 ValIleAsnTyrSerGlyValAsnAsnThrAlaTyrHisGlyTyrTrp 129
693 693
129 oArgAspPheLysLysThrAsnAlaAlaPheGlySerPheThrAspPhes 146
694AGAATTCGTGTGATATGTC 714
146 erAsnLeuIleAlaAlaAlaHisSerHisAsnIleLysValValLeuTsp 162
715 TTTTATGATGATATAGTACTTAACAGCAGCTTTGGAGATTAAGATTCA 764
163 PheAlaProAsnHisThrAsnProAlaSerSerThrAspProSerPheAl 179
765 GGATTAACAATCTTAATCTGATGGGAAATTTTATCTAAGAAATTAATG 814
179 agLAsnGlyAlaLeuTyrAsnAsnGlyThrLeuLeuLysLysTyrSerA 196
815 ATGACACCAAGAAATTCATATATCTTGTCACTTCTCTTGGCGGTTAA 864

```

196 snasprhrtaaglyleuphehishts.....AsnglyGlyThr 208
865 AGTTAACTCCACA...AATGAACGTCATTAAATATGCTGACGACG 911
209 AspheserhrtrnglusergelyltyrlysnleuTyraAspleuAl 225
912 TTCTCTGGAACAAACATGGAATAGATTAACGCTGCACCATGCCAGAA 961
225 aAspIleasnclnlnasnasnlnThrIleaspsrTyrleuysgluseri 242
962 TCAGTGTCCGGGTGTGTGTCGCTGACGCTTGCATTTGAGTCAAAA 1011
242 leglnleuTrpleuasnleuGlyValasplgylleargPheaspaIval 258
1012 GTGGAAATGCCGCGCTGACAA.....TATATGGGTATAT..... 1050
259 LysHisMetProGlnGlyTrpGlnLysSerTyrValSerSerIleTyrSe 275
1051 .....AATGTTACTTTCACA 1065
275 rserAlaasnProValptherhr 282

seq_name: SwissProt_40:RF1M_YEAST

seq_documentation_block:
ID RF1M_YEAST STANDARD; PRT; 413 AA.
AC P30775:
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Peptide chain release factor 1, mitochondrial precursor (MRF-1).
GN MRF1 OR YGL143C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP MEDLINE=93117110; PubMed=1475194;
RA Pel H.J., Maat M.J., Rep M., Grivell L.A.;
RT "The yeast nuclear gene MRF1 encodes a mitochondrial peptide chain
RT release factor and cures several mitochondrial RNA splicing
RT defects."
RL Nucleic Acids Res. 20:6339-6346(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / F11769;
RX MEDLINE=97197983; PubMed=9046099;
RA Voet M., Defoor E., Verhasselt P., Riles L., Robben J., Volckaert G.;
RT "The sequence of a nearly unclonable 22.8 kb segment on the left arm
RT chromosome VII from Saccharomyces cerevisiae reveals ARO2, RPL9A,
RT Tipl, MRF1 genes and six new open reading frames."
RL Yeast 13:177-182(1997).
CC -i- FUNCTION: MITOCHONDRIAL PEPTIDE CHAIN RELEASE FACTOR THAT DIRECTS
CC THE TERMINATION OF TRANSLATION IN RESPONSE TO THE PEPTIDE CHAIN
CC TERMINATION CODONS UAA AND UAG.
CC -i- SUBCELLULAR LOCATION: Mitochondrial.
CC -i- SIMILARITY: BELONGS TO THE PROKARYOTIC AND MITOCHONDRIAL RELEASE
CC FACTORS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X60381; CAA42932.1; -
CC EMBL: X99960; CAA68219.1; -
CC EMBL: Z72665; CAA96855.1; -
CC PIR: S28602; S28602.

```

```

DR SGD; S000311; MRF1.
DR InterPro: IPR000352; Pep_rel_factor_1.
DR Pfam: PF00472; RF-1; 1.
DR PROSITE: PS00745; RF_PROK_1; 1.
KW Protein biosynthesis; Mitochondrion; Transit peptide.
FT TRANSIT 1 413 MITOCHONDRION (POTENTIAL).
FT CHAIN 1 413 PEPTIDE CHAIN RELEASE FACTOR 1.
SQ SEQUENCE 413 AA; 46770 MW; AAMD4829748C7604 CRC64;

alignment_scores:
Quality: 95.00 Length: 307
Ratio: 0.642 Gaps: 14
Percent Similarity: 48.208 Percent Identity: 21.824

alignment_block:
US-09-839-894-9 x RF1M_YEAST
US-09-839-894-9 x RF1M_YEAST
Align seg 1/1 to: RF1M_YEAST from: 1 to: 413

268 AGCAGTTCATCGCTCAGCGGTGAACAAT.....ATACATTACAATT 311
:::|||||:::|||||:::|||||:::|||||:::|||||:::
30 ThrSerThrTrpTrpAsnSerLysSerAsnGlySerIleProHnGlnTy 46
312 TACGGAACAAAGAGTTTAATA.....AAAAGAGAGCTAC 346
:::|||||:::|||||:::|||||:::|||||:::|||||:::
46 rThrcGlnLeuSerProleuLeuValylsGlnAlaGlnLysrGlnLag 63
347 AAATTAAGAGCTATTAACAATATTATGTTCAAAAGTGTAACTGCCATCC 396
:::|||||:::|||||:::|||||:::|||||:::|||||:::
63 luleuLysAsp.....lLeuAspLysAspLeuSerCys..... 73
397 GCCTTAACACTTACCTCAGCTCATTTACCTGTAATAA..... 435
74 .....GlyIleHisPheAspValAsnLysGlnLysHsTyr 85
436 .....AACGCGCTTCAGCTGCAAGTTTATTTATTTATTTCTCTG 475
85 rAlaLysLeuSerAlaLeuThrAspTrpPheIleClnTyrLysGlnLysL 102
476 CTGCGCAACTTAACAAATTTGGCTTTGGTGGTATCTGGATCTGCTG 525
:::|||||:::|||||:::|||||:::|||||:::|||||:::
102 euAsnGlnLeuLysSerLeuGlnGlnMetIleValSerAspProSerLeu 118
526 AAGTTAAGAGTAAAGACGATATAGTAC.....ACGTATGGAAC 566
:::|||||:::|||||:::|||||:::|||||:::|||||:::
119 ArgAlaGlnLagLugLugLugLugLugLugLugLugLugLugLugLug 135
567 TTACACTATTAATATACATTTAATTAATTAATTAATTAATTAATTAAT 616
:::|||||:::|||||:::|||||:::|||||:::|||||:::
135 rThSerSerArgLeuValAsnLysLeuLeuPro..... 146
617 TATGCTTACCTCAGTCAAAAGTGACGCTGCGCTGATCTTAACCTGCT 666
:::|||||:::|||||:::|||||:::|||||:::|||||:::
147 .....ProHisProPheAlaAspLysProSerLeuLeuGlnLeuArg 160
667 CCAACTGGTGGGGC.....ACATATATGGAAGAAATCTGTGTA 707
:::|||||:::|||||:::|||||:::|||||:::|||||:::
161 ProGlyValGlyGlyIleGlnAlaMetIlePheThrIleAsnLeuAsn 177
708 TATGCTCTTTATGATGATGATATAGTACTAACAGCAGCTTTGGAGATA 757
:::|||||:::|||||:::|||||:::|||||:::|||||:::
177 pMet.....TyrIleGlyIleGlyAla...AsnTyrArgLysTrpLysTyr 191
758 GATTTCAGATTAACAATCTTAATCTGATGGAATTTATTCGAAGAA 807
:::|||||:::|||||:::|||||:::|||||:::|||||:::
191 rGlnIleSerLysAsnGlnLagLugLugLugLugLugLugLugLugLug 207
808 ATTAAGATGACACCAAGAAATTCATATCTTTCATCTTCTGCTGCTG 857
:::|||||:::|||||:::|||||:::|||||:::|||||:::
208 lLeuSerIleGlnLagLugLugLugLugLugLugLugLugLugLugLug 224
858 GGGTAAA.....AGTTAACTCCAAACAAATGGAACGT 889

```

```

224 aglyValHisArgValGlnArgIleProSerThrGluThrIleGly.... 239
890 CATTAATAATGTCGACGACCTCTCTGGAACAACACTGCAATGCAATT 939
240 .....ArgThrHisThrSerThrAla 246
940 ACAGCTGCACCATGCCAGAAATC..... 963
247 AlValValValIleProGlnIleGlySpGluSerAlaIleSeriLeAs 263
964 AGTTCGGGTGTGTGCTGCGCTGACGCTTGCATGATGCAAG 1012
263 PALATyGlnArgThrPhenylsProGlyIleArgValAspIleMet 280
1013 TGGAAATCCGAGGCTGACAAATATATGGTAA.....ATT 1050
280 rGalAsSerGlySglYgIlnHisValAsnThrThrAspSerAlaVal 296
1051 AATGTACTTTCACACCAAGT 1071
297 ArgLeuThrHisIleProSer 303
seq_name: SwissProt_40:WAPA_BACSU
seq_documentation_block:
ID WAPA_BACSU STANDARD; PRT: 2334 AA.
AC 007833;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Wall-associated protein precursor.
GN WAPA OR N176.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=93302506; PubMed=8336082;
RA Foster S.J.;
RT "Molecular analysis of three major wall-associated proteins of a
RT Bacillus subtilis 168: evidence for processing of the product of a
RT gene encoding a 258 kDa precursor two-domain ligand-binding
RT protein."
RL Mol. Microbiol. 8:299-310(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / BGSJAL;
RX MEDLINE=95219088; PubMed=7704263;
RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;
RT "Cloning and sequencing of a 29 kb region of the Bacillus subtilis
RT genome containing the hut and wapa loci."
RL Microbiology 141:337-343(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / BGSJAL;
RX MEDLINE=97124196; PubMed=8969509;
RA Yoshida K.-I., Shindo K., Sano H., Seki S., Fujimura M., Yanai N.,
RA Miwa Y., Fujita Y.;
RT "Sequencing of a 65 kb region of the Bacillus subtilis genome
RT containing the hlc and cel loci, and creation of a 177 kb contig
RT covering the gnt-sacX region."
RL Microbiology 142:313-3123(1996).
CC -1- FUNCTION: STILL UNKNOWN. NOT INVOLVED IN CELL MEMBRANE METABOLISM,
CC MOTILITY, SECRETION OR DIFFERENTIATION.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. MAY BE RELEASED
CC INTO THE MEDIUM.
CC -1- DOMAIN: HAS TWO LIGAND-BINDING DOMAINS; THE N-TERMINUS, HAS THREE
CC 101 AA REPEATS WHICH ARE RESPONSIBLE FOR CELL WALL BINDING; THE
CC C-TERMINUS CONSISTS OF TWO BLOCKS OF RESIDUES WITH A CONSERVED
CC MOTIF REPEATED 31 TIMES.

```

```

CC -1- SIMILARITY: THE REPEATED MOTIF IN THE C-TERMINUS HAS SOME
CC SIMILARITY TO THE REPEAT IN E. COLI RHS GROUP OF PROTEINS (RHS-A-D).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L05634; AAA22883.1; -
DR EMBL; D31856; BAA06656.1; -
DR EMBL; D29985; BAA06260.1; -
DR EMBL; D83026; BAA11683.1; -
DR EMBL; Z99124; CAB15959.1; -
DR PIR; S32920; S32920.
DR Subtilisin; BG10797; WAPA.
DR InterPro: IPR003305; CBD_6.
DR Pfam: PRO2018; CBD_6; 1.
KW Cell wall; Repeat; Signal; Complete proteome.
FT SIGNAL 1
FT CHAIN 29
FT DOMAIN 504
FT REPEAT 504
FT REPEAT 636
FT REPEAT 769
FT DOMAIN 1021
FT REPEAT 1021
FT REPEAT 1021
FT REPEAT 1040
FT REPEAT 1042
FT REPEAT 1063
FT REPEAT 1083
FT REPEAT 1109
FT REPEAT 1129
FT REPEAT 1150
FT REPEAT 1174
FT REPEAT 1199
FT REPEAT 1219
FT REPEAT 1238
FT REPEAT 1265
FT REPEAT 1667
FT REPEAT 1690
FT REPEAT 1711
FT REPEAT 1732
FT REPEAT 1751
FT REPEAT 1772
FT REPEAT 1795
FT REPEAT 1820
FT REPEAT 1840
FT REPEAT 1861
FT REPEAT 1887
FT REPEAT 1906
FT REPEAT 1927
FT REPEAT 1948
FT REPEAT 1969
FT REPEAT 1983
FT REPEAT 2008
FT REPEAT 2027
FT REPEAT 2047
FT REPEAT 2051
FT REPEAT 2070
FT REPEAT 2071
FT REPEAT 2093
FT REPEAT 2120
FT REPEAT 2139
SO SEQUENCE 2334 AA; 258329 MW; B75138CCD278BA3 CRC64;

```

```

alignment_scores:
  Quality: 95.00
  Ratio: 0.633
  Percent Similarity: 46.729
  Length: 321
  Gaps: 20
  Percent Identity: 22.118

```

```

alignment_block:
US-09-839-894-9 x WAPA_BACSU
Align seg 1/1 to: WAPA_BACSU from: 1 to: 2334

```

```

85 GATGAAGCATTAATAATTTTGGCCCGGTCAGAGCAAGATCTTC 134
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
260 AspergillusleuaspGluSerThrGlyValGluArgSerasp...Ly 275
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
135 CCCCACATATATATTAATTAAC.....CATATTA 166
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
275 sValSerThrLysLeuGluLysAsnGluGlyThrLeuLeuHisLeu 292
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 CAGCA.....TACAGTAAAGTCATACCTGCTAT..... 195
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
292 hValAspGluAsnThrLeuLysAspProGluValThrProValSer 308
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
196 ...GATAGGATGACTTTTATGTTGCTCTGCACATATACCTTAATG 242
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
309 IleAspProSerThrSerLeuSerValSerSerAspThrPheValMet 325
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
243 AGCATGTCCAAACGATGAGATCCTAGACGTTCATCGGTGACGGGTGAA 292
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
325 ValAlaThrProThr...AsnThrSerAlaSerSerGlnLysThrAsp 341
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
293 CAATATTAACATTAATTAATTCAGGAAAAAGAGTTTAATAAAGAGAG 342
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
341 laasn..... 342
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
343 CTACAAATTAAGGCTATAAACAATTAATTTTCAAAAGTTAACTGCC 392
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
343 .....LeuLysAlaLys..... 346
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
393 ATCCGGCCTAACACTTAACCTACGCTCATTTTAAGTAAATAAACCGCG 442
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
347 .....ValLeuLysThrGlyLys.....AspLysT 356
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
443 CTTCAGGTGCAAGTTTATATTTATATATCTGCTGGCACTAAATAAT 492
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
356 hThrThrGlyThrAsn...TyrAlaPheMetLysPheAsnAsnLeuLysPro 371
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
493 TTGCCTTTTGGTGATCTGGATGCTACTCTGAAGTTAAGATAAAG 542
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
372 IleGlnAsnMetThrValThrLysAlaThrLeuLysThrThrValAlaHis 388
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
543 AGCATATAGTACGACCTATGAGACTTACACTATAATATCTACTTAAT 592
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
388 sSerThr.....TyrGlyThr.....LysA 395
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
593 TAACTGTAAAGGAATATTCAGTATGCTTACTCTCAAGTTCAAAAGT... 639
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
395 laThrGly.....LeuThrPheAsnThrValAsnSerAsn 406
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
640 .....GACGCTGCGTCGATCTTAAGTGGCTCAACTGGTGGGCGCAC 683
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
407 TyrAspAsnAlaLysValThrThrAsnThrLysProAlaSerLysAsn.. 422
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
684 ATATATTGGAAGAAATCTGTGATATGCTTTATGATGATAT... 729
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
423 .....IleGlyLysAlaAspValHisLysGlyGlnThrAlaSerLysPv 438
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
730 .....AGTACTAACAGACGCTCTTGAGATTAAGATTT 762
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
438 alThrAlaAlaValLysSerThrAsnSerGlyGlyAlaAsnThrGlyPhe 454
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
763 CAGGATTAACAATCTTAATCTGATGGAATTTTATCTAAGAAATAA... 810
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
455 Lys...LeuHisThrAsnGlyAsnGlyLysGluThrLysLysLeuI 470
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
811 .....AATGATGACCAAGAATGTCATATCTT 841
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
470 eSerSerAlaAsnSerAlaAsnLysProThrIleGluValThrThrThr 487
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
842 TGTCACCTCTCTTGGCGTAAAGTTTAACTCAACA..... 879
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
487 lePro.....LysGlyAsnThrProThrIleLysAlaLys 498
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
880 ...AATGGAACGTCATTAAATATTGCTGACGACGCTTCTGTGAACAAA 926

```

```

499 HisAsnGlyAspSerThrGlyThrPheasp.....IleSe 510
927 CTGCAATATGAAAT 939
510 rThrLysLysVal 514
seq_name: SwissProt_40:N100_YEAST

```

```

seq_documentation_block:
ID N100_YEAST STANDARD: PRT: 959 AA.

```

```

AC 002629;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Nucleoporin NUP100/NSP100 (Nuclear pore protein NUP100/NSP100).
GN NUP100 OR NSP100 OR YKL068W OR YKL336.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93054906; PubMed=1385442;
RA Wente S.R., Rout M.P., Blobel G.;
RT "A new family of yeast nuclear pore complex proteins.";
RL J. Cell Biol. 119:705-723(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=94378724; PubMed=8091863;
RA Rasmussen S.W.;
RT "Sequence of a 20.7 kb region of yeast chromosome XI includes the
RT NUP100 gene, an open reading frame (ORF) possibly representing a
RT nucleoside diphosphate kinase gene, tRNAs for His, Val and Trp in
RT addition to seven ORFs with weak or no significant similarity to
RT known proteins.";
RL Yeast 10:569-574(1994).

```

```

CC -1- FUNCTION: ESSENTIAL COMPONENT OF NUCLEAR PORE COMPLEX.
CC NUCLEOPORINS MAY BE INVOLVED IN BOTH BINDING AND TRANSLLOCATION OF
CC THE PROTEINS DURING NUCLEOCYTOPLASMIC TRANSPORT.
CC -1- SUBCELLULAR LOCATION: Nuclear pore complex.
CC -1- DOMAIN: CONTAINS G-L-F-G REPEATS.
CC -1- SIMILARITY: BELONGS TO THE GLFG FAMILY OF NUCLEOPORINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).

```

```

CC EMBL: Z15035; CAAT8753.1;
DR EMBL: X75780; CAA53406.1;
DR EMBL: Z28068; CAA81905.1;
DR PIR: B44402; B44402.
DR PIR: S39173; S39173.
DR PIR: S44518; S44518.
DR SGD: S000151; NUP100.
DR InterPro: IPR004325; Nucleoporin_FG.
DR Pfam: PF03093; Nucleoporin_FG; 24.
KW Nuclear protein; Transport; Repeat.
FT DOMAIN 33 571
FT 29 X 6 AA APPROXIMATE REPEATS OF
SQ SEQUENCE 959 AA; 99988 MW; D3985F9901BBA51 CRC64;

```

```

alignment_scores:
Quality: 94.50 Length: 410
Ratio: 0.480 Gaps: 17
Percent Similarity: 47.073 Percent Identity: 19.512

```

alignment_block:
US-09-839-894-9 x N100_YEAST ..
Align seg 1/1 to: N100_YEAST from: 1 to: 959

```

91 AGCATTAATTAATTTTGGCCGCGTGACAGAACGATCTTCC..... 135
   ||| : : : : : ||| : : : : : ||| : : : : :
284 SeransSerGlyLeuPheGlyGlnAsnSerMetAsnSerSerThrGlnG1 300
136 .....CCCAACATATATTTA.....ATAACCATTA 163
   : : : : : ||| : : : : : ||| : : : : :
300 yValPheGlyGlnAsnAsnGlnMetGlnIleAsnGlyAsnAsnAsnA 317
164 TTACAGCATTAAGTGAAGTCACTCTGTATGATAGATGACCTTTTAA 213
   : : : : : : : : : : : : : : : : : : : : :
317 snSerLeuPheGlyLysAlaSnThrPheSerAsnSerAlaSerGly 333
214 TGTGTGCTTCTCACAATATACATTAATGAGACATGTCACACAGTGA 263
   : : : : : : : : : : : : : : : : : : : : :
334 LeuPheGlyGlnAsnAsnGlnGlnGlnGlnGlnGlnGlnGlnGln 350
264 TCCTAGCATTCATCGGTCACGGT..... 288
   | : : : : : ||| : : : : : ||| : : : : :
350 nSerGlnThrSerGlySerGlyLeuPheGlyGlnAsnAsnGlnLysG 367
289 .....GAAACAATATATACATTAACATTTACG 315
367 InProAsnThrPheThrGlnSerAsnThrGlyIleGlyLeu...PheGly 382
316 GAAAAAAGCACTTAATTAATAAAGAGAGCTACCAATTAAGCTATATAA.. 363
   : : : : : : : : : : : : : : : : : : : : :
383 GlnAsnAsnAsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 399
364 .....C 364
399 cAlaGlyThrThrGlySerLeuPheGlyGlyAsnSerSerThrGlnProA 416
365 AATATATGTTCAAAAGTGTAACTACGCCATCCGCGCTAACCTTAACATCA 414
   : : : : : ||| : : : : : ||| : : : : :
416 snSerLeuPheGlyThrThrAsnValProThrSer..... 427
415 GCTCATTTTAAGTGAATTAATAAAGCGGCTCAGTGCAAGTTATATT 464
   ||| : : : : : ||| : : : : : ||| : : : : :
428 .....AsnThrGlnSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 438
465 ATATATTTCTCTGCGCACTAAAAAATTTGCTTTGGTGTATCTGGG 514
   ||| : : : : : ||| : : : : : ||| : : : : :
439 .....GlyAlaThrLysLeuThrAsnMetProPheGlyGlyAsnProT 453
515 ATGCTACTGAAGTGAAGTAAAGACATATAGTACACCTATATGA 564
   ||| : : : : : ||| : : : : : ||| : : : : :
453 hValA.....AsnGlnSerGlySerGlyAsnSerLeuPheGly 465
565 ACTTACACTATAATATCATATTAAATTAACCTGATAAGGAAATATATCA 614
   ||| : : : : : ||| : : : : : ||| : : : : :
466 ThrLysProAlaSerThrThrGlySerLeuPheGlyAsnAsnThrAlaSe 482
615 GATATGTTACCTCACTCAAAAGT..... 639
   : : : : : : : : : : : : : : : : : : : : :
482 rThrThrValProSerThrAsnGlyLeuPheGlyAsnAsnAlaAsnAsn 499
640 .....GACGCTCGCGTGCATCTTAAC 660
499 erThrSerThrThrAsnThrGlyLeuPheGlyAlaLysProAspSerGln 515
661 TTGGCTCACT...GGTGGGCGACATAT..... 687
   : : : : : ||| : : : : : ||| : : : : :
516 SerLysProAlaLeuGlyGlyGlyLeuPheGlyAsnSerAsnSerAsnSe 532
688 .....ATTGGAAGAAAT.....TCTGTGTATA 709
   ||| : : : : : ||| : : : : : ||| : : : : :
532 rSerThrIleGlyGlnAsnLysProValPheGlyGlyThrThrGlnAsnT 549

```

```

710 TGTGCTTTATGATGATATAGTACTTAACAGCAGCTTTG..... 750
   : : : : : : : : : : : : : : : : : : : : :
549 hrGlyLeuPheGlyAlaThrGlyThrAsnSerSerAlaValGlySerThr 565
751 ..GAGATAGATTTTCAGGATTAACAATCTTAATCTGATGGAAATTTTA 797
   : : : : : ||| : : : : : ||| : : : : :
566 GlyLysLeuPheGlyGlnAsnAsnAsnThrLeuAsnValGlyThrGlnAs 582
798 TCTAAGGAAATTAATGATGACACCAAGAAATTCATATACTTTGTAC 847
   : : : : : ||| : : : : : ||| : : : : :
582 nValProValAlaAsnAsnThrThrGlnAsn..... 592
848 TTCCTGCGCGGTAAAGTTAACTCCACAATATGAAAGCATTAAT 897
   ||| ||| : : : : : ||| : : : : : ||| : : : : :
593 ..AlaLeuLeuGlyThrThrAlaValPro.....SerLeuGln 604
898 ATTGCTGACGAGCTTCTGTGAAACAATCGAATATGATTAACAGCTGT 947
   ||| : : : : : ||| : : : : : ||| : : : : :
605 GlnAlaProVal.....ThrAsnGlnGlnLeuPheSerLysI 617
948 CACCATGCCA...GAATCAGTGTCCGCTGTGTGTGCTGACGCTT 994
   : : : : : ||| : : : : : ||| : : : : :
617 eSerIleProAsnSerIleThrAsnProValLysAlaThrThrSerLysV 634
995 TGCATTTGATGCAAAAGTGAAGAAATCCGAGCTGACCAATATATGAGGT 1044
   : : : : : ||| : : : : : ||| : : : : :
634 AlaAsnAlaAspMetLysArgAsnSerSerLeuThrSerAlaThrArgLeu 650
1045 AATATTAATGTTACTTTCACACCAAGTAGT 1074
   : : : : : ||| : : : : : ||| : : : : :
651 AlaProLysProLeuPheAlaProSerSer 660

```

seq_name: SwissProt_40:LEU2_BUCRP

```

seq_documentation_block:
ID LEU2_BUCRP STANDARD; PRT; 471 AA.
AC P48573;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE 3-isopropylmalate dehydratase large subunit (EC 4.2.1.33)
DE (Isopropylmalate isomerase) (Alpha-IPM isomerase) (IPMI).
GN LEUC.
OS Buchnera aphidicola (subsp. Rhopalosiphum padi).
OG Plasmid PRPE.
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=98793;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95333198; PubMed=7608990;
RA Bracho A.M., Martinez-Torres D., Moya A., Latorre A.;
RT "Discovery and molecular characterization of a plasmid localized in
RT Buchnera sp. bacterial endosymbiont of the aphid Rhopalosiphum
RT padi."
RL J. Mol. Evol. 41:67-73(1995).
CC -1- CATALYTIC ACTIVITY: 3-ISOPROPYLMALATE = 2-ISOPROPYLMALATE +
CC H(2)O (ALSO CATALYSES 2-ISOPROPYLMALATE + H(2)O = 3-HYDROXY-
CC 4-METHYL-3-CARBOXYPENTANONE).
CC -1- PATHWAY: SECOND STEP IN LEUCINE BIOSYNTHESIS.
CC -1- SUBUNIT: CONSISTS OF TWO DIFFERENT SUBUNITS: LEUC AND LEUD.
CC -1- SIMILARITY: BELONGS TO THE ACONITASE/IPM ISOMERASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: X71612; CA50617.1; -.
CC InterPro: IPR001030; Aconitase.
CC Pfam: PF00330; aconitase; 1.

```


FT CARBOHYD 98 98 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 164 164 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 412 AA; 45948 MW; E420D3530D1A04D9 CRC64;

Alignment_scores:

Quality: 93.50 Length: 305
 Ratio: 0.645 Gaps: 14
 Percent Similarity: 47.341 Percent Identity: 18.361

Alignment_block:

US-09-839-894-9 x THBG_SHEEP ..

Align seg 1/1 to: THBG_SHEEP from: 1 to: 412

```

97 ACTAATATTTTGGCCCGGTGACGAGACGAACTTCCCAACATTA 146
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
155 ThrGluValPheSerThrAspPheSerAsnValSerAlaIleGlnGlu 171
   ||| |||:|||||:|||||:|||||:|||||:|||||:|||||:
147 TATATTAATACCATTAATACAGCATACGTAAGT..... 183
   :|| |||:|||||:|||||:|||||:|||||:|||||:|||||:
171 uile...AsnSerHisValGluArgGlnThrLysGlyIleValGlyL 187
   :||| |||:|||||:|||||:|||||:|||||:|||||:|||||:
184 .....CATACTCTGATGATAGATGACTTTTGA 213
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
187 euIleGlnAspLeuLysProAsnThrIleThrValLeuValAsnTrpLeu 203
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
214 TGTTCGTCCTTCACATATACACTTAATGAGACANTGCCACAGTGA 263
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
204 CysPheLysAla...GlnTrpAlaAsnProPheAspProSerLysThrG 219
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
264 TCCTAGCAGTCATCGTCGACGCGTGAACAATAATTAACATTACA 313
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
219 uGIuGIySerSerPheLeuValAspLysThr...ThrValGln.... 233
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
314 CGGAAAAAGAGTTTAAATAAAGAGCTACAAATTAAGGCTATAAA 363
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
234 .....ValProMetMetHisGlnValAspGlnTrpTyr 244
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
364 CAATTATTTGTTCAAAAGTGTAACTGCCACATCGGCTAACACTTA 413
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
245 HisLeuValAspThrGluLeuAsnCys.....ThrVa 255
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
414 AGCTCATTTTAACGTGTAATAAAGCGGCTTCAGGTCAACTTATAT 463
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
255 IleuGlnMetAspTyrSerLysAsnAla.....LeuAlaLeuPhe 269
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
464 TATATATTCCTGCTGCGAAGCTAAATAATTCCTTTGGTGTATCTG 513
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
269 aIleuProLysGluGlyGlnMetGluGlyVal..... 279
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
514 GATGCTACTCTGAAGTAAAGATAAAGACATATAGAGACCTATAG 563
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
280 GluGlyAlaMetSerSerLysIleLeuLysLysTrpAsn..... 292
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
564 AACTTACACTATAATATCATTATAATTAAGTAAAGGGAATATTC 613
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
293 .....ArgLeuLeuGlnLysGlyTrpValA 301
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
614 AGATATGTTTACCTCAGTTCAAAAGTGAACGCTCGGCTGATCTTA 663
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
301 snLeuPheValProLysPheSerIleSerAlaThrTyrAspLeu.... 315
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
664 CGTCCAACTGTGTGG..... 678
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
316 .....GlyGlyIleLeuLeuLysMetGlyIleGlnAspAlaPheAl 329
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
679 .....GGCACATATATTTG 691
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
329 aAspAsnAlaAspPheSerGlyLeuThrLysAspAsnGlyLeuLysVal 346
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
692 GAAGAAATTCCTGTGATATGCTTTTATGATGATATAGTACTTAAC 741
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

346 erAsnValAlaHisLysAlaMetPheTyrIleGlyGluLysGlyThrGlu 362
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
742 AGCTCTTGGAGATAGATTTTCAGATACAAATCTTAATCT..... 783
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
363 AlaIleProGluValArgPheLeu...AsnGlnProGluThrThrLeu 378
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
784 .....GATGGAAATTTATCTAAGCAAAATTAATG 814
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
378 uHisProIleIleGlnPheAspArgSerPheLeuLeuIleLeuGlu 395
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
815 ATGACACCAAGAAATTCATATACTTTGTCACTTCTCTTGGCGGTAA 864
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
395 yAsnThrArg.....SerIleLeuPheLeuGlyLys 405
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
865 AGTTTAAGTCCACA 879
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
406 ValValAspProThr 410
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

This Page Blank (uspto)

Tue Jul 2 09:01:36 2002

OM of: US-09-839-894-9 to: SPREMBL_19:* out_format : pfs

Date: Jul 1, 2002 11:22 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL-frame+map.model -DEV+1p
 -Q/cgcn2.1/USPTO.spool/US09839894/runat_01072002.111122.17530/app-query.fasta.1.1160
 -DB-SPREMBL_19 -QFMT-fastan -SUFFIX-rspt -GAPOP-12.000
 -GAPEXT-4.000 -MINMATCH-0.100 -LOOPEL-0.000 -LOOPEXT-0.000
 -GAPOP-4.500 -GAPEXT-0.050 -XGAPOP-10.000 -XGAPEXT-0.500
 -FGAPOP-6.000 -FGAPEXT-7.000 -YGAPOP-10.000 -YGAPEXT-0.500
 -DELOP-6.000 -DEEXT-7.000 -START-1 -MATRIX-biosum62
 -TRANS-human0.cdi -LIST-45 -DOCALLIG=200 -THR_SCORE-pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE-LOCAL -OUTFMT=pts
 -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER-US09839894.ecgnl.1.327 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-839-894-9
 Query length: 1086
 Database: SPREMBL_19:*
 Database sequences: 562222
 Database length: 172994929
 Search time (sec): 128.580000

score_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
SP_bacteria:Q9367	1886.00	3025.46	9.9e-161	361	Q9367 escherichia coli. minor
SP_bacteria:Q47115	949.50	1517.95	9.2e-77	363	Q47115 escherichia coli. coid
SP_bacteria:Q47119	930.50	1487.34	4.6e-75	364	Q47119 escherichia coli. coid
SP_bacteria:Q9X051	275.50	433.12	2.5e-16	359	Q9X051 salmonella typhi. and sa
SP_bacteria:P71132	117.50	171.46	0.0393	847	P71132 chlamydomophila abortus. f
SP_bacteria:P71133	111.00	161.01	0.1305	846	P71133 chlamydomophila abortus. f
SP_bacteria:P94750	109.00	149.07	0.2506	2349	P94750 escherichia coli. invas
SP_bacteria:P76347	108.00	148.95	0.2510	2383	P76347 escherichia coli. o2383
SP_bacteria:Q9RLA0	108.00	155.92	0.2803	872	Q9RLA0 rickettsia typhi. dna pc
SP_bacteria:Q25579	108.00	146.83	0.3103	2529	Q25579 helicobacter pylori. c6
SP_fungi:Q96VW2	105.50	157.94	0.4391	430	Q96VW2 blumeria graminis. extra
SP_bacteria:Q9GFL1	104.50	146.00	0.6057	1441	Q9GFL1 lactococcus lactis (sub
SP_plant:Q9A0U5	104.00	150.19	0.6352	803	Q9A0U5 oryza sativa (rice). hyf
SP_bacteria:Q9R8S7	103.50	150.67	0.6342	691	Q9R8S7 dehnococcus radiodurans.
SP_archaeap:Q97206	103.50	145.47	0.7356	1269	Q97206 sulfolobus solfataricus
SP_bacteria:Q9ZK89	103.50	140.04	0.7817	2399	Q9ZK89 helicobacter pylori. j95
SP_invertebrate:Q95PH7	102.50	141.19	0.9317	1736	Q95PH7 dictyostelium discoid
SP_bacteria:Q9L800	102.50	137.27	0.9734	2747	Q9L800 aeromonas salmonicida.
SP_archaeap:Q9Y1M6	102.00	148.24	0.9464	692	Q9Y1M6 sulfolobus solfataricus.
SP_bacteria:Q46962	100.00	141.76	1.48	1014	Q46962 escherichia coli. cni1
SP_plant:Q9LX60	100.00	137.09	1.56	1752	Q9LX60 arabidopsis thaliana (m
SP_bacteria:Q9CKZ7	99.50	138.83	1.68	1300	Q9CKZ7 pasteurella multocida.
SP_bacteria:Q92C16	99.50	126.32	1.94	5627	Q92C16 pseudomonas aeruginosa.
SP_bacteria:Q48508	99.00	146.97	1.69	456	Q48508 lactobacillus acidophilu
SP_bacteria:Q9ZC16	99.00	142.15	1.78	802	Q9ZC16 listeria innocua. pheny
SP_bacteria:P77922	99.00	141.76	1.79	839	P77922 chlamydia psittaci (chla
SP_bacteria:Q9PHY4	99.00	141.59	1.79	355	Q9PHY4 campylobacter jejuni. p
SP_bacteria:Q47888	98.50	147.39	1.85	395	Q47888 fremyella diplophilum (c
SP_bacteria:Q9ZHI0	98.50	125.86	2.35	4919	Q9ZHI0 haemophilus ducreyi. la
SP_bacteria:Q93QW7	98.00	148.02	2.02	334	Q93QW7 bacteroides fragilis. m
SP_bacteria:Q9ZK42	98.00	143.21	2.13	587	Q9ZK42 listeria innocua. lln061
SP_bacteria:Q47106	98.00	139.19	2.24	1014	Q47106 escherichia coli. cyto
SP_fvitus:Q66961	97.50	139.19	2.45	855	Q66961 feline immunodeficiency
SP_archaeap:Q92812	97.00	143.79	2.55	1755	Q92812 methanotermobacter (m
SP_plant:Q81310	97.00	138.92	2.69	454	Q81310 arabidopsis thaliana (m
SP_bacteria:Q93440	97.00	143.79	2.69	803	Q93440 paenibacillus sp. kcc88
SP_bacteria:Q46342	96.50	128.89	3.31	2364	Q46342 clostridium sorbelloi.
SP_bacteria:Q9F0F4	96.00	143.13	3.10	406	Q9F0F4 mycoplasma mycoides sub
SP_plant:Q93ZM5	96.00	142.64	3.12	430	Q93ZM5 arabidopsis thaliana (m
SP_organelle:Q21375	96.00	135.50	3.38	992	Q21375 pichia kluyveri (yeast)

SP_plant:Q80973 + 95.50 135.58 3.71 895 1 Q80973 arabidopsis thaliana
 SP_archaeap:Q9HJ37 + 95.50 128.37 4.02 2081 1 Q9HJ37 thermoplasma acidoph
 SP_bacteria:Q9K541 + 95.00 117.66 3.56 198 1 Q9K541 mycobacterium paratub
 SP_bacteria:Q9K6M4 + 95.00 138.76 3.94 561 1 Q9K6M4 bacillus halodurans.
 SP_bacteria:Q93G68 + 95.00 135.05 4.10 867 1 Q93G68 escherichia coli. ush

seq_name: SP_bacteria:Q93G67

seq_documentation_block:

AC Q93G67 PRELIMINARY; PRT; 361 AA.
 ID Q93G67;
 AC Q93G67;
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE MINOR PILIN PROTEIN CSAE.
 GN CSAE.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=E1181A;
 RA Altboum Z.D., Levine M.M., Galen J.E., Barry E.M.;
 RT "Isolation and characterization of EPEC C54 fimbriae encoding genes,
 RT and their expression in Shigella flexneri 2a guana strain CVD 1204.";
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF296132; MAF97137.1;
 SQ SEQUENCE 361 AA; 40102 MW; 2E1E74ABDB0EB353 CRC64;

alignment_scores:

Quality: 1886.00 Length: 361
 Ratio: 5.224 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-839-894-9 x Q93G67 ..
 Align seg 1/1 to: Q93G67 from: 1 to: 361

1 ATGAATTAAGATTTTATTTATTTTACATGTTTCTCTGACTTTT 50
 |||||||
 1 MetanilsyleuphellepetherleupheneSerValleuph 17
 51 TACATTGCTGATGCGAGATTAATCCGAGATGAAGCATTAAC 100
 |||||||
 17 ethrphelaValserAlaAsplylleproglyaspIuserIethra 34
 101 AATTTTGGCCGCGTACAGACGAATCTTCCCAATATATATA 150
 |||||||
 34 snlIepheIlyproArpaspagaSngIuserSerProlyshIsnIle 50
 151 TTAATTAACATATTACAGCATACAGTGAAGTCACTTATGATAG 200
 |||||||
 51 LeuadnAsnHsIlethAlaItyrSerGluSerHsIthIleuTyAspAr 67
 67 gneIthrphelauSyleuserSerHsIasnIthleuasnIyAlaCysP 84
 251 CAACAGTGAAGATCTGACAGTCACTGCGTACGGGTGAACAAATTA 300
 |||||||
 84 IofHserCluasnProserSerSerValserCllyIthrshnIle 100
 301 ACATTAACATTTACGAAAAAGAGTTTAATAAAGAGACTCAAAAT 350
 |||||||
 101 ThrleuGlnpHerthCluIySargSerleuIeIySargGluIeGlnI 117
 351 TAAAGCTATAAACAATTTATTTGTTCAAAAGTGTAACTGCGCGGCC 400
 |||||||
 117 elySelyIyIyIySgInleuIeuphelysSerValasnCysProserGlyL 134

401 TAACACTTAAGCTGATTTTAACTGTAATAAAGCGGCTTCAGCT 450
 |||||||
 134 eutrhleuansserAlahisPheasnCyasnLysasnAlaLaserGly 150
 |||||||
 451 GCAAGTTATATATATATATCCGTGGCGCACTAAATAATTTGCCCTT 500
 |||||||
 151 AlaserleuYrleuYrileProAlaGlyleuLeuLysasnLeuProPh 167
 |||||||
 501 TGGTGTATCTGGAGTCTACTCTGAAGTTAAGATAAAGACGATATA 550
 |||||||
 167 eGlyGlyletrpaspAlaThrleuLysleuAlaValLysArgArglyTS 184
 |||||||
 551 GTGAGACCTATGAACTTACACTATAATATACATTAATAATTAACGAT 600
 |||||||
 184 eGluThrYrGlyThrYrThrileasnileThrileLysleuThrasp 200
 |||||||
 601 AAGGGAATATTCAGATATGTTACCTACCTCAAAAGTGCAGCTCGGT 650
 |||||||
 201 LysGlyasnileGlniletrPleuProGlnPheLysSeraspAlaArgVa 217
 |||||||
 651 CGATCTTAAGTGGTCCCACTGGTGGGCGACATATATTTGGAAGAATT 700
 |||||||
 217 LaspLeuasnleuArgProThrGlyGlyGlyThrYrileGlyArgasNS 234
 |||||||
 701 CTGTGATATGCTTTTATGATGATATAGTACTTAACAGCAGCTCTTG 750
 |||||||
 234 eValaspMetCysPheYrAspGlyYrSerThrAsnSerSerleu 250
 |||||||
 751 GAGTATAGATTTCAAGATACAAATCTTAATCTGATGGAAATTTTATCT 800
 |||||||
 251 GluileArgPheGlnaspasnProLysSeraspGlyLysPheYrLe 267
 |||||||
 801 AAGAAATATATGATGACACCAAGAAATGCAATATCTTGTGCACATC 850
 |||||||
 267 uArgLysileasnspAspThrLysGluileAlaYrThrleuSerleuL 284
 |||||||
 851 TCTTGGCGGGTAAAGTTTAACTCCAAACAATGGAACGTCATTAATAT 900
 |||||||
 284 eulAlaGlyLysSerleuThrProThrAsnGlyThrSerleuasnile 300
 |||||||
 901 GCTACGAGCTCTCTCGAACAACAACGGAATGAATTAACGCTGTAC 950
 |||||||
 301 AlaaspAlaLaserleuGluThrAsnTrpAsnArgileThrAlaValTh 317
 |||||||
 951 CATGCCAGAAATCAGTCTCCGCTGTTGTTGCTGAGCTGAGCTTGCAT 1000
 |||||||
 317 rMetProGluileSerValProValleuGlyTrpProGlyArgleuGln 334
 |||||||
 1001 TGGATGCAAAAAGTGAATAATCCGAGGCTGACATATATGGTAATAT 1050
 |||||||
 334 euAspAlaLysValGluasnProGluAlaGlyGlnYrMetGlyasnile 350
 |||||||
 1051 AATGTACTTTCACACCAAGTAGTCAAACTC 1083
 |||||||
 351 AsnValThrPheThrProSerSerGlnThrleu 361

seq_name: sp_bacteria:047115

seq_documentation_block:
 ID 047115 PRELIMINARY; PRT; 363 AA.
 AC 047115;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE COOP PRECURSOR.
 GN COOP.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94344028; PubMed=7915003;

RA Froehlich B., Karakashian A., Melsen L.R., Wakefield J.C., Scott J.R.;
 RT "Cooc and Cood are required for assembly of CSI pil.";
 RL Mol. Microbiol. 12:387-401(1994).
 DR EMBL: X76908: CAA54230.1
 DR InterPro: IPR001412; TRNA-synt.L.
 DR PROSITE: PS00178; AA-TRNA_LIGASE_I; UNKNOWN_1.
 KW Signal.
 FT SIGNAL. 1 18 POTENTIAL.
 FT CHAIN 19 363 COOD.
 SO SEQUENCE 363 AA: 40139 MW: AA9488AADD5F72 CRC64;

alignment_scores:
 Quality: 949.50 Length: 371
 Ratio: 3.230 Gaps: 11
 Percent Similarity: 79.245 Percent Identity: 53.100

alignment_block:
 US-09-839-894-9 x 047115

Align seg 1/1 to: 047115 from: 1 to: 363

7 AAGATTTTATTTATTTTATTTTACATTTGTTTCTCTTCACTATTTTACATT 56
 |||
 2 LysLysilePheilePheleu.....SerilelePheSerAl 14
 |||
 57 TGCCTGATNCGCGACATATAAATCCCGGAGATGA.....AGCATACTA 100
 |||
 14 aValValSerAlaGlyArgYrProGluThrThValGlyasnleuThrL 31
 |||
 101 ATATTTT...GGCCCGCGT...GACAGAAAGCATCTTCCCAACAT 144
 ::|||
 31 YSerPheGlnAlaProArgleuaspArgSerValGlnSerProiletyr 47
 |||
 145 AATATATTAATTAACCATATTACAGCATACAGTGAAGTCACTACTCTGA 194
 |||
 48 AsnilePheThrAsnHisValAlaGlyYrSerleuSerHisSerleuY 64
 |||
 195 TGATAGATGACTTTTATGTTGTTGCTCTCAATACATACACTTAATGAG 244
 |||
 64 rAspArgileValPheleuGlyThrSerSerSerAsnProValasnly 81
 |||
 245 CATGCCAAGCAGTGAAGATCTGACAGTTCATGCGTACG...GCTGAA 291
 |||
 81 lacYsProThrle.....GlyThrserGlyValGlnYrGlyThr 94
 |||
 292 ACAATATATACATTACATTTACGGAAGAAAGATTATTAATAAAGA 341
 |||
 95 ThrThrileThrleuGlnPheThrGluLysArgSerleuileuLysArgas 111
 |||
 342 GCTCAATATTAAGGCTATATAACCAATATGTTGTCAAAAGTGAAGTCC 391
 |||
 111 nileasnleuAlaGlyasnLysLysProiletrPheLysnleuSerYsa 128
 |||
 392 CATCCGCG.....CTAACCTTAAGCTCACTTTTAACTGAATATAA 435
 |||
 128 sPheSerAsnleuMetValleuasnSerLysSertrpSerCysGlyAla 144
 |||
 436 AAGCG...GCTTCAAGTGCAGTTTATATTTATATATCTGCTGGCGA 482
 |||
 145 HisGlyasnAlaasnGlyThrleuLeuasnleuYrileProAlaGlyL 161
 |||
 483 ACTAATAAATTTGCTTTGTTGTTGATCTGAGATCTACTCTGAAGTTAA 532
 |||
 161 uileasnLysleuProPheGlyGlyiletrPheAlaThrleuileuAla 178
 |||
 533 GAGTAAAGACGATATAGTACAGC.....TATGACACTTAC 570
 |||
 178 rgleuSer...ArgTyrGlyValSerSerThrHisTyrGlyAsnYr 193
 |||
 571 ACTATTAATATACATTTAATTAACGATTAAGGAAATATTCGATATG 620
 |||
 194 ThrValasnileThrValaspLeuThrAspLysGlyasnileGlnValTr 210

```

621 GTTACTCAGTCAAAAGTGCAGCTGCGATCTTAACTTGCAGCA 670
||||| ||| |||: |||||||||||||||
210 PLEPFGIYPhenHisSerAsnProAlaValAspLeuAsnLeuArgProI 227
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
671 CTGGTGGGGACATATATTTGAAGAATTCGTGTATGATGCTTTAT 720
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
227 leglyAsnTyrLysTyrSerGlySerAsnSerLeuAspMetCysPheTyr 243
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
721 GATGGATATATGCTACACACAGCTCTTTGGAGATTAAGATTACAGTAA 770
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
244 AspGlyTyrSerThrAsnSerAspSerMetValIleLysPheGlnAspAs 260
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
771 CAATCCCAATCTGATGGGAAATTTTATCTAGAGAAATTAATGATACAA 820
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
260 PAsnProThrAsnSerSerGlyTyrAsnLeuTyrLysIleGly...GlyT 276
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
821 CCAAGAAATTCATATACCTTGTCTGCTCTGGCGGTTAAAGTTTA 870
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
276 hrgIuLysLeuProTyrAlaValSerLeuMetGlyGluLysIlePhe 292
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
871 ACATCCAAATGGAACGCTCATTAATATTTGCTGACGACGCTTCTGGA 920
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
293 TyrProValAsnGlyInsSerPheThrIleAsnAspSerSerValLeuG 309
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
921 ACAAATGGAATAGATTAATACAGCTGCACCATGCCAGAAATCAGTTC 970
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
309 uThAsnTyrAsnArgValAlaThrAlaValAlaMetProGluValAsnValP 326
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
971 CGGTGTGTGTGGCTGACGCTTTCGCAATTCGATGCAAAAGTGGAAT 1020
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
326 rovalleuGlyTyrProAlaArgLeuLeuLeuAsnAlaAspValAsnAla 342
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1021 CCCGAGCTGACCAATATATGGAATATTAATGTTACTTACACCAAG 1070
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
343 ProAspAlaGlyGlnTyrSerGlyGlnIleTyrIleThrPheThrProse 359
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1071 TAGCAACACTC 1083
| : : : : : |||
359 rValGluAsnLeu 363
| : : : : : |||
seq_name: sp_bacteria:047119

```

```

seq_documentation_block:
ID 047119 PRELIMINARY; PRT: 364 AA.
AC 047119;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE COTD PRECURSOR.
GN COTD.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C91F-6;
RX MEDLINE=96071908; PubMed=7591145;
RA "Groehlich B.J., Karakashian A., Sakellaris H., Scott J.R.;
RT "Genes for CS2 pilli of enterotoxigenic Escherichia coli and their
RT interchangeability with those for CS1 pilli."
RL Infect. Immun. 63:4849-4856(1995).
KW Signal.
FT CHAIN 1 18 POTENTIAL.
FT SIGNAL 19 364 COTD.
SQ SEQUENCE 364 AA: 40979 MW: EC7AC738A0E195B6 CRC64;

```

```

alignment_scores:
Quality: 930.50 Length: 366
Ratio: 3.209 Gaps: 3

```

```

Percent Similarity: 79.235 Percent Identity: 48.361
alignment_block:
US-09-839-894-9 x 047119
Align seg 1/1 to: 047119 from: 1 to: 364
1 ATGAAATAGATTTATTTATTTTACATGTTTTCTTCAGTACCTTT 50
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1 MetLysValIlePheValLeuSerMetPheLeuGlySerGlnValTyr 17
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
51 TACATTTGCTGATCGGCAGATAAATTCGCGAGTAAAGCAATACAT 100
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
17 rGlyInsSerThrIleThrAsnValGluAlaGlySerIleAsnLysThrG 34
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
34 InsrIleGlyProIleAspArgSerAlaAlaIleSerTyrProAlaHis 50
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
101 ATATTTTGGCCCGCTGACAGAAC.....GAATCTCCCAACAT 144
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
34 InsrIleGlyProIleAspArgSerAlaAlaIleSerTyrProAlaHis 50
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
145 AATATATTAATACATATTACAGCATACAGTAAAGTCATACCTGTA 194
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
51 TyrIlePheHisGluHisValAlaGlyTyrAsnLysAspHisSerLeuPh 67
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
195 TGATAGATGACTTTTATGTTTGTCTCTCACATACACTTATAGAG 244
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
67 eAspArgMetThrPheLeuGlySerMetSerThrAspAlaSerLysGly 84
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
245 CATGCTCCACACAGTACAGATTCAGAGTTCATCGGTACCGGTAAACA 294
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
84 lacysProThrGlyGluAsnSerLysSer.....GlnGlyGluThr 98
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
295 AATATACATTACATTTACGGAAAAAGATTATTAATAAGAGAGCT 344
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
99 AsnIleLysLeuIlePheThrGluLysLysSerLeuAlaArgLysThrLe 115
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
345 ACAATTAAGGCTATTAACATATATGTTCAAAAGGTTAAAGCTGC... 390
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
115 uAsnLeuLysGlyTyrLysArgPheLeuTyrGluSerAspArgCysIleH 132
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
391 ....CCATCGGCTTAACACTTACCTGACCTCATTTTAACTGAATAAA 435
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
132 lstrValAspLysMetAsnLeuAsnSerHisThrValLysCysValGly 148
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
436 AACCGGCTTCAGGTGCAAGTTATATTTATTTCTGCTGCGCAACT 485
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
149 SerPheThrArgGlyValAspPheThrLeuTyrIleProGlnGlyIuI 165
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
486 AAAAAATTTGCTTTGCTGATCTGGAGTACGACCTGAGTAAAGAG 535
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
165 eAspGlyLeuLeuThrGlyGlyIleThrGluAlaThrLeuGluLeuArgV 182
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
536 TAAAGACGATATATAGTACACCTATGGAACCTTACATATAATATACAT 585
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
182 allysArgHisTyrAspLysAsnHisGlyThrTyrLysValAsnIleThr 198
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
586 ATTAAATTAACGATTAAGGAATATTCAGATATGTTACTCTCAAGTTCAA 635
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
199 ValAspLeuThrAspLysGlyAsnIleGlnValIleThrProLysPheH 215
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
636 AAGGAGCGTCCGCTGATCTTAACCTGCGCAACGAGGCGGACAT 685
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
215 sSerAspProArgIleAspLeuAsnLeuArgProGlnGlyAsnGlyLysT 232
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
686 ATATTTGGAAGAAATTCGTGATATGCTTTTATGATGATATATAGTACT 735
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
232 yserGlySerAsnValLeuGlnMetCysLeuTyrAspGlyTyrSerThr 248
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
736 AACAGAGCTCTTTGAGATTAAGATTTCAGAGATTAACATCTTAATCTGA 785
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
249 HisSerGlnSerIleGlnMetArgPheGlnAspSerGlnThrGlyAs 265
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
786 TGGGAATTTTATCTAGGAAATTAATGATGACACCAAGAAATTCAT 835
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

265  nmsnGluYrAenLeuIleYltsYThrGlyGluProLeuYlSylsLeuProt 282
836  A+ACTTTCGACCTCTCTCTGGGGGATAAAGTTTAACTTCACCAAAATGA 885
282  YLysLeuSerLeuLeuLeuGlyGlyIleArgGluPhePheProAsnSlnly 298
886  ACGTCATTAAATATTCGTGCAGCCACTCTCTGGAACAAACTGAAATAG 935
299  GluAlaPheThrIleAsnAspHisSerSerLeuPheIleAsnTyrPnaR 315
936  AATTACACTCTCCACCAATGCCAGAAATACAGTTCCTCCGCTGTGTCTGC 985
315  GtLeysSerValSerLeuProGlnIleSerIleProValLeuCyStrp 332
986  CTGGACGTTTGCACATTCGATGCACAAAGTGGAATAACCCGAGGCTGGACA 1035
332  roAlAsnLeuThrPheMetSerGluLeuAsnAsnProGluValGlyIu 348
1036  TATATGGGTAATATTAAGTTACTTTTCACACCAAGTACTGCAACACTC 1083
349  TyrSerGlyIleLeuAsnValIleThrPheTrpProSerSerSerLeu 364

```

seq_name: sp_bacteria:Q9XDS1

```
seq_documentation_block:
ID      Q9XDS1      PRELIMINARY;      PRT;      359 AA.
```

DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE TSD PROTEIN.
GN TSD OR TCFD.
OS Salmonella typhi, and
OS Salmonella enterica.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=601, 28901;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=GIU10007;
RA Hashimoto Y.;
RT "Identification of a putative fimbrial operon of Salmonella typhi.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
LN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.enterica; STRAIN=RKS 3333;
RX MEDLINE=99348391; Pubmed=10417651;
RA Lofthagen A., Advani A., Sukupolvi S., Pfeifer J.D., Normark S.,
RA Jordehan S.;
RT "Multiple insertions of fimbrial operons correlate with the evolution
RT of Salmonella serovars responsible for human disease.";
RL Mol. Microbiol. 33:612-622(1999).
DR EMBL: AB029403; BAA82272.1; -
DR EMBL: AJ242964; CAB51577.1; -
SO SEQUENCE 359 AA; 39741 MW; BB88A4E9A62052CE CRC64;

```
alignment_scores:
  Quality: 275.50
  Ratio: 1.514
  Percent Similarity: 60.066
  Length: 303
  Gaps: 11
  Percent Identity: 29.703
```

```
alignment_block:
US-09-839-894-9 x Q9XDS1
```

Align seg 1/1 to: Q9XDS1 from: 1 to: 359

208 TTTTATGTTTGCTCCTCCACATTACACTTAATGGAGCATGTCCACCAG 257
:::||||| |||||:::|||||:::
79 TPVtYcYargserfsmnrgasnngluansnglujYalAcyscglnugjrh 95
:::|||||:::|||||:::|||||:::
256 TGAGCAATCCTAGCAGTTTCATCGGTGACGCCGTGAACAACAATATAACATTAC 307
:::|||||:::|||||:::|||||:::

[illegible]

779AsnProIleasnAspHisThrValThrPheAlaValLeuSer	792
178	GAAGTCATCACTCTGTATGATAGATGACTTTTATGTGTCCTCTCA	227	
	::::::::::::::	:::::	
793	GlySerAlaThrSerPheAsnSerGlnAsn	
ThraIleLy	
805	Ly	805	
228	CATACACTTAATGGAGCATGTCACCAACGAGACATCTTGACAGCTTCAT	277	
	::: :::::::::: :::: ::::	:::	
805	SThrAspValasnGlyLeuAla...ThrPheAspLeuSerSerLeuSerG	821	
278	CGGTCCGGCTGAACAAATATACATTACATTTACGGAAGAAAGAAAGT	327	
821	IncluspasnThrValGluValThrLeuGluasnGlyValLysLeuThr	837	
	:::	
328	TTTAATA.....	333	
838	LeuIleValSerPheValGlyPserSerThrAlaGlnValAspLeuG	854	
334	
	
854	lySerSerLysasnGluValValAlaAspGlyAsnAspSerValThrMet	871	
374	TCAAAAGTTTAACTGCCATCCGCCCTTAACACTTAACCTGACAGCTCATTT	423	
	:::	:::	
871	hValAlaThrValArgAspAlaValGlyAsnLeuLeuAsnAspAlaMetVal	887	
424	AACTGTAAATAAACGGCGCTCAGCTGCAAGTTATATTATATATATTC	473	
	::: ::: :::	:::	
888	ThrPheAsnValAlaSerAlaLeuAlaValLysLeuSerGlnThrGluValAs	904	
474	TGCTGGCGAAGCTAAAAAATTTGCTTTGGTGATCTGGAGTGGCTACTG	523	
	::::::		
904	nSerHisAsp.....	GlyLeuAlaThrAlaThrL	
524	TGAAGTTAAGACTAAAAAGACATATAGTGAGACCTATGAGAACTTAACT	573	
914	euThr.....	..SerLeuLysasnGlyAspIly	
	
574	ATTAATATCATATTAAATTAAGTATGATAAGGAATATTTCAGATATGTT	623	
	:::~::~:	:::	
924	ValThrAlaSerValSerSerLysSerGlnAlaAsnGlnGlnValAsnPh	940	
624	AACTCAGTTCAAAAGTGAACGCTGGCGTGAGTCTTAACCTGCGTCCAACTG	673	
	:::~::~:	:::	
940	eIleGlyAspGlnSerThrAlaAlaLeuThrLeuSerVal...ProSerG	956	
674	GTGGGGGCACATATATATGSAAGAAATTCGTGATATATGCTTTATATAT	723	
956	lyAspIleThrVal.....	960	
724	GGATATAGTACTAAGCAGCAGC.....	TCCTTGGAGATTAAGATT	
	:::	:::::	
961ThrAsnThrAlaProGlnInlyrMetThrAlaThrLeuGlnAs	
	
762	TCAGATTAACAATTCCTAAATCTGATGGAAATTTATCTAAGAGAAATAA	811	
	:::~::~:	:::	
974	plyAsnasnGlyasnProLeuLysAspLysGlnThrPheSerValProA	991	
812	ATGATGTACACCAAAAGAAATGCATATACTTTGTCACTTCTCTGGCGGGT	861	
991	snAsp.....	..ValAlaSerLysPheSerIleSerAsnLysGly	
	
862	AAAGTTTAACTCAACAAATGAGACGTCATTA.....	894	
	:::	:::	
1004	LysGlyMetThrAspSerasnGlyValAlaIleAlaSerLeuThrGlyTh	1020	
895ATATTG	
	
1020	rleuAlaGlyThrHisMetIleMetAlaArgLeuAlaSerSerAsnValS	1037	
	
902	CTGAGCAGACTTCTCTGGAAACAACCTGGAATTCAGATTACAGCTGTACC	951	
	:::~::~:	:::	
1037	eTrAspAlaInProMetThrPheValAlaLysLysAspArgAlaValAl	1053	

```

995 ATGCCGAAATACAGTTCGCGGTGCGTGGCCGTGA...CGTT 995
::: ::: ||| ::: |||
1054 VALLEUglnThrSerLysAlaGluIleIleLysnclYlValspcluth 1070
||| ||| |||::: |||::: |||::: |||:::
1070 rThleuThrAlaThrValLysAspPro...SerAnHisProValAlaG 1086
1046 ATATTAACTGTTACTTTCACA 1065
::|||::|||::|||::|||
1086 lYlEthrValAsnPherh 1092

seq_name: sp.Bacteriap: p76347

seq_documentation block:
ID P76347 PRELIMINARY; PRT; 2383 AA.
AC P76347;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE 02383.
GN B1978.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RA Blatter F.R., Plunkett G. IIL, Mayhew G.F., Perna N.T., Glasner F.D.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 2118-2383 FROM N.A.
RC STRAIN=K12;
RA Alba H., Baba T., Fujita K., Hayashi K., Honjo A., Horuchi T.,
RA Ikemoto K., Inda T., Isono K., Isono S., Itoh T., Kanai K., Kasai H.,
RA Kasahimoto K., Kim S., Kimura S., Kitagawa M., Kitakawa M., Makino K.,
RA Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y.,
RA Nishimoto H., Nishio Y., Oshima T., Saito N., Sampei G., Seki Y.,
RA Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano M.,
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE000289; AAC75042.1; -.
DR EMBL; D90837; BAAL15800.1; -.
DR InterPro: IPR003344; Big_1.
DR InterPro: IPR003535; Intimin.
DR InterPro: IPR002482; LysM.
DR InterPro: IPR000601; PKD_domain.
DR Pfam; PF02369; Big_1; 13.
DR PRINTS; PRO1369; INTIMIN.
DR SMART; SM00257; LysM; 1.
DR SMART; SM00089; PKD; 4.
KW Complete proteome.
SQ
SEQUENCE 2383 AA; 251391 MW; 129975F4C64225C CRC64;

alignment_scores:
Quality: 109.00 Length: 390
Ratio: 0.565 Gaps: 15
Percent Similarity: 49.487 Percent Identity: 20.000

alignment_block:
US-09-839-894-9 x P76347 ..

Align seg 1/1 to: P76347 from: 1 to: 2383

37 TCTTCAGTACTTTTACATTTCGTATCGGCACATAAATATCCCGAGA 86
||||: ::| |||::: |||::: |||::: |||:::
762 SerAlaLysIleAlaThrLeuSerAlaSerAsnnglyValLeuAlaAs 778
87 TGAAGCACTACTAATATTTTGGCCCGCGAGCAGACAAGCATCTCC 136
||||: ::| |||::: |||::: |||::: |||:::
778 ngLusnAlaAlaAsnThrValSerValAsnValAlaAspGluGlySer. 794

```

137	CCAAACATTAATATTATTAATTAACCAATTAACGACATAC.....	AGT	177
795AsnProIleAsnAspHisIsthValIthPheIaValLeuSer		808
178	GAAGATCATACCTCGTATGATAGATGACCTTTTATGATGTTGCTCTTCA		227
809	GlySerAlaThrSerPheAsnAsnGlnAsn.....	ThIaIaLy	821
228	CATATACCTTAATGGAGCATGTCTCCAAACCAAGTGAAGATCTGACACTGAT		277
821	SthIrsPValAsnGlyLeuAla.....	ThPheSpleuIlySerSerIySG	837
278	CGTCGACGCGTAAACAAATATATACATTACATTTACGGAAAAAGAGT		327
837	InGlnAspAsnThrValGluValAlThrIleGluAsnGlyValIySelsInhr		853
328	TTTATA.....		333
854	LeuIleValSerPheValGlyIySperSerThrIaGlnValAspLeuGly		870
334AAACAGAGCTACCAATTAATAAGCCATATAACAAATTA...TUGT		373
870	AllySerIySAsnGlyValValAlAlAspGlyAsnAspSerValIthrMet		887
374	TCAAAAGTGTTAACGTGCCCATCCGGCTCAACACTTAACTCAGACCTCATTT		423
887	hcrAlaThrValArgSpAlaIySgIyAsnLeuLeuAsnAspAlaMetVal		903
424	AACTGTAAATAAAAACCGCGCTTCACAGTCGACGATTATATATATATATCC		473
904	ThrPheAsnValAsnSerAlaGlnAlaIySLeuSerGlnIthrGluValAs		920
474	TGCTGGCGAACAATAAAATTTGCTTTGGTGGTATCTCGAGTGTACTGC		523
920	nserIthAsp.....	GlyIleAlaThrAlaThrL	930
524	TGCAAGTTAAGATAAAAGACGATTAATAGTACACTTAATGGAACCTTAACT		573
930	euthr.....	SerIleuIySAsnGlyAspIyArg	939
574	ATTAATATCTATTAAATTAATTAATCTGATTAAGGAATAATTCACATATGTT		623
940	ValIthAlaSerValSerSerIySerGlnAlaAsnGlnGlnValAsnPh		956
624	ACCTCAGTTCAAAAGTACGACCTCGCGTCGATCTTAACCTTGCGTCCACTG		673
956	eIleGlyAspGlnSerIthAlaAlaIeuthrLeuSerVal...ProSerG		972
674	GTCGGGGCACATATATNTGSAAGAAATTCGTGTGATATGCGTTTATGAT		723
972	IyAspIleThrVal.....		976
724	GGATATAGTACATACAGACAGC.....	TCTTTGGAGATAGATT	761
977	ThrAsnThrAlaProGlnIyrmethrAlaIthIleuGlnAs	990
762	TGAGGATTAACAATCTCTTAATTCGATGGGCAATTTATCTAAGAAAAATAA		811
990	plysAsnGlyAsnProLeuIySAspIleGlnIleIthrPheSerValProA		1007
812	ATGATGCACACCAAGAAATTCGATTAACCTTGTCTACTCTCTGGCGGGT		861
1007	snAsp.....	ValAlaSerIySperSerIleSerAsnGlyIy	1019
862	AAAAAGTTTAAGTCCAAACAAATGAAGACGTCATTA.....		894
1020	IySgIyMetThrAspSerAsnGlyValAlaIleAlaSerLeuThrGlyTh		1036
895	AATATG	901
1036	rIeuaIaGlyThrHisMetIleMetAlaArgLeuAlaAsnSerAsnValS		1053

```
902 CTGACGAGCTCTCTGTGAAACAAACTGATAGATTACAGTGTGACC 951
    :|||||
    :|||||
1053 eRAspAlaInPrOmEtThrPhEValAlAspLyAspArgAlaValVal 1069
    :|||||
    :|||||
952 ATGCCAATAACAGTGTCCGGTGTGTGTGGCTGGA.....CGTTT 995
    :|:|:|
    :|:|:|
1070 ValLeuInThrSerLySAlaGluIleIleGlyAsnGlyValAspGluTh 1086
    :|:|:|
    :|:|:|
996 GCATGTGATGACAAAGTGGAAATCCCGAGCGACGAATATATGGGTA 1045
    :|:|:|
    :|:|:|
1086 rThrLeuThrAlaThrValLySAspPro...SerAsnHISproValAlaG 1102
    :|:|:|
    :|:|:|
1102 LyIleThrValAsnPhetHr 1108
    :|:|:|
    :|:|:|
seq_name: sp_bacteria:Q9RLA0

seq_documentation_block:
ID Q9RLA0 PRELIMINARY; PRT; 872 AA.
AC Q9RLA0:
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-OCT-2001 (TReMBLrel. 18, Last annotation update)
DE DNA POLYMERASE I.
GN POLA.
OS Rickettsia typhi.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiinae; Rickettsia.
OX NCBI_TaxId=785;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WILMINGTON:
RX MEDLINE=99416441; PubMed=10486973;
RA Anderson J.O., Anderson S.G.E.;
RT "Genome degradation is an ongoing process in Rickettsia.";
RL Mol. Biol. Evol. 16:1178-1191(1999).
DR EMBL: AJ238757; CAB56108.1; -.
DR HSSP: P05582; IKPD.
DR InterPro: IPR002421; 5_3_exonuclease.
DR InterPro: IPR002288; DNA_pol_I.
DR InterPro: IPR001098; DNA_pol_A.
DR InterPro: IPR000513; Exo_N.I.
DR InterPro: IPR003584; HHH.2.
DR Pfam: PF01367; 5_3_exonuclease; 1.
DR Pfam: PF02739; 5_3_exonuc_N; 1.
DR Pfam: PF00476; DNA_pol_A; 1.
DR PRINTS: PR00868; DNAPOLI.
DR SMART: SM00475; 53EXOC; 1.
DR SMART: SM00279; HHH2; 1.
DR SMART: SM00482; POLAC; 1.
DR PROSITE: PS00447; DNA_POLYMERASE_A; 1.
DR SEQUENCE 872 AA; 99309 MW; BC2EBE3490879A08 CRC64;
```

```
alignment_scores:
    Quality: 108.00      Length: 378
    Ratio: 0.607        Gaps: 15
    Percent Similarity: 47.090      Percent Identity: 20.370
```

```
alignment_block:
US-09-839-894-9 x Q9RLA0 ..
```

```
Align seg 1/1 to: Q9RLA0 from: 1 to: 872
```

```
4 AATAAGATTTTATTATTATTATTTACATGTTTCTCTCTCAGTACTTTTAC 53
    |||||
    :|:|:|
383 AsnLySerIleLySAlaGluIleThrPhEValAlAspLyAspArgAlaVal 399
    :|:|:|
    :|:|:|
54 ATTGCTGTATCGGACAGATAAATTCGCCGAGATGAAGCAT.....96
    :|:|:|
    :|:|:|
399 eTyAlaAsnGlnSerHisLySAlaGluIleThrAlaIleGluAspLeuLeu 416
```

```
97 .....ACTAATATTTTGGCCCG 114
    :|:|:|
    :|:|:|
416 eTgInTyThrLeuSerAlaGlyLeuValGlnLySAsnLeuPhAla... 431
    :|:|:|
    :|:|:|
115 CGTGACAGGAACGATCTCCGCCAACATATATATTAATACATAT 164
    :|:|:|
    :|:|:|
432 .....GluThrLeuThrLySAspAsnIleIleAsnGluSerAl 444
    :|:|:|
    :|:|:|
165 TACAGCATACAGTGAAGTCACTACTGTATGATAGATGATGCTTTTAT 214
    :|:|:|
    :|:|:|
444 aLySAlleValIleAsnPhelIleSerLeuTyf...LySAlInThrLeuEua 460
    :|:|:|
    :|:|:|
215 GTTGTCTCTTCACAAATACACTTAAT.....GGA 243
    :|:|:|
    :|:|:|
460 lAlaGlnLySAsnLySAlaPhaArgLeuTyfArgGluIleAspLeuPro 476
    :|:|:|
    :|:|:|
244 GCATGTCCCAACCACTGAGAAATCTGACATTCATCGGTACGGTGAAC 293
    :|:|:|
    :|:|:|
477 ThrcySpheIleLeuAspLySmetGluLySAlleGlyIleLySValAspAl 493
    :|:|:|
    :|:|:|
294 AAATATACATTAATTAATTAAGAAAA.....AGAAGTTAATAAAA 337
    :|:|:|
    :|:|:|
493 aAsnTyfLeuHISGlnLeuSerAspGluPhcGlyThrcGluIleLeuLyS 510
    :|:|:|
    :|:|:|
338 GAGAGCTACAAAT..... 351
    :|:|:|
    :|:|:|
510 lGluGlnGluIlePhaAlaLeuSerGlyThrcLySPhaAsnIleAlaSer 526
    :|:|:|
    :|:|:|
352 ...AAGGCTATAACATTAATTTGTCAAAAGTGTAACTGCCATCCG 398
    :|:|:|
    :|:|:|
527 GlnLySglnLeuSerGluIleLeuPhcLySmetGlnLeuProSerG 543
    :|:|:|
    :|:|:|
399 CCTAATCACTTACACGCTCATTTTAAGTAAATTAACCAACGGCTCAG 448
    :|:|:|
    :|:|:|
543 yAsnThrcLeuAlaLySThrcSerTyfSerThLyS..... 555
    :|:|:|
    :|:|:|
449 GTGCAAGTTTATTTATTTATTCCTGCTGCGCACTAAATAATTTGCT 498
    :|:|:|
    :|:|:|
556 .....AlaGlyIleLeuLySLeuSer 563
    :|:|:|
    :|:|:|
499 TTTCGTGATCTGGATGCTACTCTGAAGTTAAGA..... 534
    :|:|:|
    :|:|:|
564 GluAspGlyTyfHISAlaThrcLeuLeuArgTyfArgGlnLeuTh 580
    :|:|:|
    :|:|:|
535 ....GTAAAAGACATATAGTAGACCTATGGAATCTTACCTTAATA 580
    :|:|:|
    :|:|:|
580 rLySLeuLySAsnThrcTyfThrcAspSerLeuProLySglnIleAsn 597
    :|:|:|
    :|:|:|
581 TCACATTAATAATTA.....ACTGATTAAGGA 606
    :|:|:|
    :|:|:|
597 lThrcLySArgIleHISThrcThrcPhcLeuGlnThrcSerThThrcGly 613
    :|:|:|
    :|:|:|
607 AATAATGAGATATGTTACTCACTTCAAAAGTACGCTGCGTCAATCT 656
    :|:|:|
    :|:|:|
614 ArgLeuSerSerGlnGluProAsnLeuGln.....AsnI 625
    :|:|:|
    :|:|:|
657 TAACTGCGTCCAACTGCTGGGGCACATATATTTGAAGAAATCTGTG 706
    :|:|:|
    :|:|:|
625 eProThrcArgSerSerAspGly.....AsnLySAla 636
    :|:|:|
    :|:|:|
707 ATATGCTGCTTTAT.....GATGATAT.....AGTACTAAGAC 741
    :|:|:|
    :|:|:|
636 rGlnAlaPhaPhelIleAlaGluAspGlyTyfLySLeuIleSerAlaAsp 652
    :|:|:|
    :|:|:|
742 AGCTCTTGGAGATTAAGATT..... 762
    :|:|:|
    :|:|:|
653 SerGlnIleGluLeuArgIleLeuSerHisIleAlaAsnValAspVal 669
    :|:|:|
    :|:|:|
763 .....CAGATACAAATCTTAATCTGATGGA 790
    :|:|:|
    :|:|:|
669 uLySglnAlaPhaPhelIleAsnLySglnAspIleHISThrcInThrcLacys 686
    :|:|:|
    :|:|:|
791 AATTTTATCTAAGAAATAAATGATGACACCAAGAA..... 828
```

```

686  lnlephasnleuglnlyshsgluclerhserglnlnsargarglys 702
      ::::: |||:::||||
829  .....ATGCATATACCTTGTCACCTCTCTGGC 857
      ||| ||| ::::
703  AlAlysAlAlleasnphcyllylellelyrglyleseralphcylle 719
      ||| ||| ::::
858  GCGTAAGATTAAACGCCAACAAATGGAAGCTCA 891
      :::::|||||
719  uAlAlysGlnleuAsnValSerAsnGlyTyrAla 730
      :::::|||||

```

seq_name: sp_bacteriap:025579

```

seq_documentation_block:
ID      025579          PRELIMINARY;          PRT;      2529 AA.
AC      025579;
DT      01-JAN-1998 (TrEMBLrel. 05, Created)
DT      01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT      01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE      TOXIN-LIKE OUTER MEMBRANE PROTEIN.
GN      HP0922.
OS      Helicobacter pylori (Campylobacter pylori).
OC      Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC      Helicobacter.
OX      NCBI_Taxid=210;
       [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=26695 / ATCC 700392;
RX      MEDLINE=97394467; PubMed=9252185;
RA      Tomb J.-F., White O., Keriavage A.R., Clayton R.A., Sutton G.G.,
RA      Fleischmann R.D., Ketchum K.A., Klein H.-P., Gill S., Dougherty B.A.,
RA      Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA      Loftus B., Richardson D., Dodson R., Khalak H.G., Glöck A.,
RA      McEneaney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA      Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA      Cotton M.D., Meldrum J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA      Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA      Venter J.C.;
RT      "The complete genome sequence of the gastric pathogen Helicobacter
RT      pylori."
RL      Nature 388:539-547(1997).
DR      EMBL; AE000602; AAD07969.1; -.
DR      TIGR; HP0922; -.
DR      InterPro: IPR004311; VACA.
DR      Pfam; PF03077; VACA; 3.
KW      Hypothetical protein; Complete proteome.
SQ      SEQUENCE      2529 AA; 274561 MW; 4440882E8644472BC CRC64;

```

alignment_scores:

Quality:	108.00	Length:	374
Ratio:	0.568	Gaps:	22
Percent Similarity:	50.802	Percent Identity:	22.460

```
alignment_block:
```

Align seg 1/1 to: 025579 from: 1 to: 2529

253 ACCAGGACGAATCCTACGCATTCCTCGCCGGTGAACAACAATAATTAAC 302
|||||::: |||||::: |||||::: |||||::: ||
977 ThrSerAsnPhesnAlaThrThgInLeuIeuGlyAsnThrAsnPheTh 993

303 ATTACCATTTACGGAATAAAGAATTATA..... 333
|||| :::::::::::|||::: ||:::
993 rIeu..... SerSerGIserIeuLeuAsnPheasnGlyAspThrT 1007

334AAAAGAGCGCTACCAATTAAGCGCTATAACAACTATTATG 372
::: ::::: ||::: ||::: ||||
1007 hrLeugInAsnAsnAlaasnlIerThrGlyAsnLyserGIlnAlaIa 1023

373 TTCMAA..... AGTGTTAACGTGCCCATCCGGCGCTAACCTT.. AA 410
||||| ::::::::::: |||||::: |||||::: ||

1024	PhelysAsnSerLeuThrLeuSpasnsnSerAsnLeuSerLeuAspAs	1040
411	CTCAGCTACTTTAACTGATATAAAAACGGCGCT.....TCAGGTGCA	454
1040	nginSerValLeuAsnAlaAsnAsnThrSerAlaPheAsnSngInAla	1057
455	GPTTATATATATATATATATTCCTGCT...GGCGAACTAAAAAATTTGCCATT	501
1057	erLeuAsnIleTyrsnGlySerGlnAlaThrPheAsnSerLeuPheP	1073
502	GGGTATCTGGGAGACTCTACTGTGAAGTAAAGTAAAAACGATATAG	551
1074	AsnGly.....GlyThrLeuSerLeuAsnAlaSerLeuSlyLeuAs	1087
552	TGAGACCTTGGAACTTACACTATATATATACATTTAAATTAACGTAT	600
1087	nAlaSerAsnAlaSerPheSerAsnAsnThrThrLeuAsnLeuSpas	1104
601AAGCGAAATATTCACGATA	618
1104	erValLeuSerAlaSerAsnThrSerSerLeuAsnAlaAsnIle.....	1118
619	TGGTACCTACAGTCAAAAGTAGCCTCGCGCATCTTTAACTTGCGTCC	668
1119AsnPhcGlnGlyAlaSerGlnAlaAspHe.....	1128
669	AACGTGTGGGGCACATATATATTTGGAAGAAATTCGTGTATGCTTTT	718
1129GlyGlyAsnThrIleIleAspThrAlaSerPheAsn.....Phe	1142
719	ATGATGGATATAGTACTACACCACTCTTGG.....GAGATA	756
1142	spSerAlaSerSerLeuAsnPheAsnAsnLeuThrAlaAsnGlyAlaLeu	1158
757	AGATTTCACGATACAACTCTAA.....	780
1159	AsnPhcAsnGlyTyThrProSerLeuThrIlysalAlaLeuMetSerValSe	1175
781TCGTATGGAGAAATTTTATCTAAGAAATTA	811
1175	rgIyGlnPheValLeuGlyAsnAsnGlyAspIleAsnLeuSerAspIle	1192
812	AT.....GATGACACCAAGAAATTCGATATACTTTGTCACTTCTC	852
1192	snIlePheAspAsnIleThrIlySerValThrTyf.....AsnIleLeu	1206
853	TTGGCGGGTAAAAAGTTTAACCTACCAAAATGGAACGTCA.....	891
1207	AsnAlaGlnIlyGlyIleThrGlyIleSerGlyAlaAsnGlyTyfGlyIly	1223
892TTAAATATGCTGAGCGACACTCTCTGGAACA	925
1223	sIleLeuPheTyfGlyMetIlyIleGlnAsnAlaThrTyfSerAspAsn	1240
926	AC.....TGGATAGCATTTACAGCTGTACCATCCAGAAATC	963
1240	snAsnIleGlnThrTyfSerPheIleAsnProLeuAsnSerSerGlnIle	1256
964AGTCTCCGGTGG.....	978
1257	IleGlnIleSerIleIlyAsnGlyAspLeuThrIleGlyAlaLeuAsnAs	1273
979TGTTGGCTGAGCGTTGTC	997
1273	nProAsnSerAlaSerAsnThrIlePheAsnIleAlaProIleuLeuTyf	1290
998	AATGTGATCAAAAGTGAATAATCC.....	1023
1290	snTyfGlnAlaSerIlyGlnAsnProThrGlyTyfSerTyfAspTyfSer	1306
1024GAGCTGGACAATAT.....ATGGTATATATATGTTACTTT	1061
1307	AspAsnGlnAlaGlyTyfThrTyfIleuThrSerAsnIleIlyGlyLeuPh	1323

OC Ehrhartoideae; Oryzeae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Bueli C.R., Yuan Q., Ouyang S., Moffat K.S., Hill J.N., Burr P.C.,
 RA Hsiao J., Zisman V., Pal G., Bowman C.L., Fujii C.Y., Vanaken S.E.,
 RA Bowman C.L., Craven B., Utterback T.R., Khalak H., Feldblum T.V.,
 RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
 RT "Oryza sativa chromosome 3 BAC OSJNBa0040E01 genomic sequence."
 RT Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC079887; AAK16187.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 803 AA: 89209 MW: E119841096A8BDAC CRC64;

alignment_scores:
 Quality: 104.00 Length: 322
 Ratio: 0.717 Gaps: 10
 Percent Similarity: 45.031 Percent Identity: 20.186

alignment block:

US-09-839-894-9 x Q9AUV5 ..

Align seg 1/1 to: Q9AUV5 from: 1 to: 803

```

97 ACTAATATTTT...GGCCCGCGTGACAGACGACATCTTCCCAACA 143
|||||
406 ThrAsnIleuAspGlyGlnSerSerLeuAsnIleuThrIleSerH 422
144 T.....AATATAT 151
422 SAspGlyLeuValAlaThrAsnGlyValIlePrlGlyHisArgAspValM 439
152 TAAATTAACAT...ATTACAGCATACAGTAAAGTCACTGTATGAT 198
|||||
439 etGlnAsnHsProIeuGlyValIleGlySerSerHisAlaArgHisPro 455
199 AGATGACCTTTTATGTGTCTTCTGCACATACACTTAATGAGCATG 248
|||||
456 GlnIleuIleuProMetIleCysThrSerHisValIleuLeuProPro 472
249 TCCAACCCAGAGATCTAGACATTCATCGTCGCGGCGGAACAATA 298
|||||
472 OAsnIleuSerAsnProAsnSerSerPheValArgGlyThrIleuLeu 489
299 TAACATTACAAATTACGAAAAAGATTAAATAAAGAGACCTACAA 348
|||||
489 yslYsAlaGluIleuMetGlyAlaIleGlySerThrThrGluAsnTyrLeuAsp 505
349 ATTTAAAGCCTATAACAATATTGTTCAAAAGTGTAACTGCCATCCG 398
|||||
506 LeuAspAsnSerGluArg.....LysGlyIleIleIleCysProLys 519
399 CCTAACCTTAACCTCAGCTCATTTTAACTTAATTAATAAAGCGCTCAG 448
|||||
519 rSerGluGlnAsnValAlaGluAsnIleCysAsnIleYsAlaIleIuTyrG 536
449 GTGCAGCTTATATTTATATATTCCTGCGCAACTA..... 486
|||||
536 IncysAsnAspTyrSerGlnIleValAlcysAsnGlnIleThrValLeu 552
487 .....AAAAATTGCGCTTTGGTGATC.....TGCGA 515
|||||
553 ProLeuAsnIleuValSerHisPheGlyLeuProThrIleYsLysPheAs 569
516 TGCTACTCTAGTTAGATAGATAAAGACGATATAGTACGACCTATGAA 565
|||||
569 pGlyLysLeuValSerArgGlnIleYsLysArgAlaThr..... 583
566 CTTACACTATTAATATCATATTAACTGAAGGAAATATTTGAG 615
584 .....SerLeuLeu 586

```

```

616 ATATGTTTACCTCAGTTCAAAAGTGAC...GCTGCGCGCATCTTAACCT 662
|||
587 SerThrPheIleAlaGlnValMetSerGlyCysSerIleThrHisIleThr 603
663 GCGTCCAGCTGTTGGGGCAGCATATATTGGAAGAAATTCCTGTGATATGT 712
|||||
603 glyProGluIleuAspIlePheIleHisAlaIleThrArgArgLeuValGluYsV 620
713 GCTTTTATGATGATATAGTACTATACAGCAGCTCTTTGGACATTAAGATT 762
|||||
620 alaPheAlaGluAsnMetLysThrIleYsAsnSerThrPhe..... 632
763 CAGGATTAACATCTTAATCTGATGGGAATTTATCTAAGAAATATATAA 812
633 .....ValIse 634
813 TGATGACACCAAGAAATTCGATATACCTTTGTCATTCTTTGGCGGGTA 862
|||||
634 rGlnAlaGlnIleYsArgLeuAlaPheThrThrIleLeuMetGlnTyrIleL 651
863 AAAGTTTAACTCCACAAATGGAACGTCATTAATATATGCTGACGCACT 912
|||||
651 euProValIleuProAspArgLeuAlaIleAlaIleAsp..... 665
913 TCCTGGAACCAACGTGAATAGAAATTACAGCTGTCACCAATCCAGAAAT 962
|||||
666 SerGlySulThrIleValIleTyrArgThrSerArgLeuAlaIleuProAspAl 682
963 CAGTGTCCGCGTGTG 978
682 aPheAsnProAlaIle 687

```

seq_name: sp_bacteriap.Q9R2S7

```

seq_documentation_block:
ID Q9R2S7 PRELIMINARY; PRT: 691 AA.
AC Q9R2S7:
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE HYPOTHETICAL 69.7 KDA PROTEIN.
GN DRB0037.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RI.
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans RI."
RT Science 286:1571-1577 (1999).
DR EMBL; AE001826; AAF12628.1; -.
DR TIGR; DRB0037; -.
DR InterPro; IPR002048; EF-hand.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Hypothetical protein; Plasmid; Complete proteome.
SQ SEQUENCE 691 AA: 69702 MW: 7441126A23C022E CRC64;

```

alignment_scores:
 Quality: 103.50 Length: 328

```

Ratio: 0.676      Gaps: 15
Percent Similarity: 46.646      Percent Identity: 21.341

alignment_block:
US-09-839-894-9 x Q9RS7 ..

Align seg 1/1 to: Q9RS7 from: 1 to: 691

82 GGAGATGAAGCATATCATATATTTTGGCCCGGTACAGAAACGATC 1111
||||: ||||| ||| :: |||
218 GlySerProThrLeuThrAsnThrValValProSerSerProAsnProLe 2344
132 TTCCCCCAACATATATATTAATAACCATATTACAGCATACAGTGAAA 1010
||| ::
234 uProProThrGlyThr..... 2399
182 GTCAATCTCTGTATGATAGATGACATTTTATGTGTGTCACAAAT 2311
||||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
240 .....ThrThrProGly 2433
232 ACACTTAATGAGCATGTCCAAACGAGAGAACTCAGAGTTCAATCGT 2810
||| ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
244 ThrGlyThrGlyProValGlySerGlyAsnGlyProIaGluSerThrGI 2606
282 CAGCGGTGAACAAATATA...ACAATTACATTTACGGAAAAAAGAGTT 3282
260 yProIaTyThrGlyIleGlySerThrSerGlyThrGlnProProValY 2777
329 TAAATAAAAGAGAGCTACCAAAATTTAAAGCGCTTAACACATATTGTCAAA 3787
||||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
277 alValaSPaLSeRcIySncIInIyralaTyrProIySlaaSPaLaGlu 2933
379 AGGTGTAACGTCCATCCGCCCTACACCTTAACGTACGCTCATTTATCTG 4282
||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
294 ThrValaSn...ProAspSerValThrMetThrThrIleThrAsnPr 3099
429 TAAATAAAAGCGGGCTGTAGGTGCAAGTTATATTATATATCTGCTG 478
||||: ||||| ||||| ||||| ||||| ||||| ||||| |||||
309 OnSnProThrIaalaThr.....TyrGluLeuValPro... 3200
479 GCGAAGCTAAAAATTTGGCTTTGGTGTATCTGGATGCTACTCTGAG 5282
||||: ||||| ||||| ||||| ||||| ||||| ||||| |||||
321 ..AspLeuSerGlyLeuProIaGly..... 3282
529 TTAAGAGTAAAAGACGATATATGTGAGACCTTAGGAATTACACTATAAA 578
328 ..... 3282
579 TATCACTTAAATTAATCACTGAT..AAGCGAAATATCAATATGCTTAC 6252
||||: ||||| ||||| ||||| ||||| ||||| ||||| |||||
329 ValThrValThrPheThrAspIaalaSnGlyAsnPro.....LeuP 3422
626 CTCACGTTCAAAAGTGACCGCTCGCTGCATCTTAACCTGCGCCCAACTGG 6752
||||: ||||| ||||| ||||| ||||| ||||| ||||| |||||
342 roAspThrAspGlyAsnGlyArgProGluValThrIaGlyProAsnGly 3582
676 GGG...GGCACATATATTGGAAGAAATTCGTTGATATGCTTTATGA 7222
||| ::||| ||||| ||||| ||||| ||||| ||||| |||||
359 GlyThrIaIaThrTy.....ArgValValValThrTyThrProAs 3711
723 TGGATATAGTACTACACGACGCTCTTTGGAGATAAGATTACAG..... 7652
||||: ||||| ||||| ||||| ||||| ||||| ||||| |||||
371 pThrGluSerIaIaIaValaValaIaGlyProIleArgIleProValaGly 3882
766 ..GATAAACAATCCTAATCTGATGGAAATTTATCTAAGGAAATTAAT 8133
||||: ||||| ||||| ||||| ||||| ||||| ||||| |||||
388 alaSPaGlyAsn.....Arg 3922
814 GATGACACCAAAATGTCATATACTTGTCACTTCTTGGCGGGTAA 8633
||| :: ||| ||||| ||||| ||||| ||||| ||||| |||||
393 AspGlyIleValaSPaLaIaThrValThrTyAsnValLeuLeuSerAsnLe 4099
864 AAGTTTAACTCCAACAATGAGACGTCATTAATATTTGCTGACGCAAGTT 9133

```

```

      ::::  ||||| ||||| ::::: |||:::
409 ulyshegclysanthrinsnglythrinalaleuglyvalseraspvalpro 426
      :::| |
914 CTTGTGAACAACATCGAATAGAAATTACAGCTGCACCATCGCAGAAATC 963
      :::| |
426 alThrArgThr.....|.....ValThrGlnProSerVal 435

964 AGTGTTCGGGTGTGTGTGGCCCTGGAGCGTTGCAATTTGGATGCAAACT 1013
      :::: ||::: ::||| |||
436 ThrThrAlaValVal...PheProMetaspLeuasp..... 448

1014 GGAAATCCCGAGCGCTGCAGCAATTTATGAGTAAAT 1047
449 .....GlyAlaIyrAspGlyasn 454

seq_name: sp_archaeap:Q97206

seq_documentation block:
ID Q97206 PRELIMINARY; PRT; 1269 AA.
AC Q97206;
DT 01-OCT-2001 (TREMBLrel. 18; Created)
DT 01-OCT-2001 (TREMBLrel. 18; Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18; Last annotation update)
DE PROTEASE RELATED PROTEIN.
CN SS01141.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Avey M.J., Chan-Welher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
RA Helkams de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Garrelts R.A., Ragan M.A., Sengen C.W., Van der Oost J.,
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus p2."
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR EMBL: AEO06730; AAK41392.1; -.
DR InterPro: IPR000209; Peptidase_S8.
DR PROSITE: PS00138; SUBTILASE_SER; UNKNOWN.1.
KW Complete proteome.
SO SEQUENCE 1269 AA; 138027 MW; 87C8F05CCF83DA84 CRC64;

alignment_scores:
      Quality: 103.50 Length: 290
      Ratio: 0.802 Gaps: 15
Percent Similarity: 44.483 Percent Identity: 20.345

alignment_block:
US-09-839-894-9 x Q97206 ..

Align seg 1/1 to: Q97206 from: 1 to: 1269

226 CAACTACACTTAATGAGACA..TGTCACAACCGTAGAATCCTACAG 272
||| |||||:::|||||::: |||::: |||||:::
790 HSIIErHrIIEgIuGlySerIIElleProProValasnuLeuProSerIn 806

273 TTCA.....TCGGTCAGCGGTGAACAATAATATCACTTACATTTA 313
::: ::::: |||||::: |||||::: |||||:::
806 rThrheGlnspAlaleuGlnGly...ThrAsnIIErHrAlAlaLysIeuV 822

314 CGGAAAAAAGAGTTTAATAAAGAGACGACTACAAATTAAGGCTATATA 363
::: |||||::: |||||::: |||||::: |||||:::
822 alserSerAsnGlyValValIIElsnGlu..... 831

364 CAATTATGTTCAAAAGCTTAAGTCCCAATCCGCGCTACACTTAAGTC 413
::: ||||| ||| |||||::: |||||::: |||||:::
832 .....AlaAsnIIEPro.....IeuSerProAsnG 840

```



```
414 AGCTCATTTTAACTAATATAAAACGGGCTCAGTGAAGTTTAAAT 463
: : : : :
840 yIleTyRPhnegly.....TyrL 846
|||||
464 TATATATTCCTGCTGGCAACTAAAAATTTGCCTTTGGTGTATCTGG 513
|||||
846 eutyRllePro.....LysAsnThrPro..SerGlyLeuTyr 857
|||||
514 GATGCTACTCTGAAGTTAAGATAAAAAGACATATAGTGAACCTATGG 563
: : : : :
858 AsnValleu.....PhealathTyRty 866
|||||
564 AACTACATATAATATCATATTAATTAATTAAGGAATATATTC 613
: : : : :
866 rSerTyRThrleuAsnThrThrIleArg..... 875
|||||
614 AGATATGGTTACCTCAGTTCAAAAGTAGCGCTCGCATTTAACTTG 663
: : : : :
875 ..... 875
664 CGTCCACACTGGTGGGCGACATATATTTGAAGAATTCGTTGATATGTG 713
||| ||| |||
876 .....GlyPheTyRtyGlyGlnIle..... 882
|||||
714 CTTTATGATGATATAGTACTACAGCAGCTCTTGAGATAGATTC 763
||| : : : : :
883 .....TyrValSerAsnGlnalathrIleSerValLyserv 895
|||||
764 AGGATAACAATCTAATCTGATGGAATAATTTATCTAAGAAATAAAT 813
: : : : :
895 alAsnTyRlAlaPhnegluginThrValPheIleTyRlAlaAsnIleThr 911
: : : : :
814 GATGACACCAAGAAATTCATATCTTGTCACTTCTCTGGCGGATA 863
: : : : :
912 AsnGlyThrAsnGlnIleLysPhe..... 919
|||||
864 AAGTTAACTCCACAAATGGAACGTCATTAATATGCTGACGACGCTT 913
|||
920 .....GlyMetPheSerAlaThrValTyRProserS 930
|||||
914 CTCGGAACAACACTGAATAGATATACGCTGCACATGCCAGAAATC 963
|||||
930 erleuSerPheAsnTyRThrThrIleSerSerIle.....Ile 942
|||||
964 AGTGTTCGGGTGTGTGTGGCCTGGACGTTTGCAATGGATGCAAAAGT 1013
: : : : :
943 GlnIleProleu.....Trp..... 947
|||||
1014 GGAATATCCGAGAGCTGGAATATATGGTATATTAATGTTACTTTCA 1063
|||||
948 .TyRAsnProLysIleGlnIleTyRglutPgluginAsnPheThrLeuProserA 964
: : : : :
1064 CACCAAGTACTCAACACTC 1083
: : : : :
964 lathSerAlaGlnAsnLeu 970
```

7

This Page Blank (uspto)